universität wien

DIPLOMARBEIT

Titel der Diplomarbeit

Schrödinger's Equation as Newton's Law of Motion

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag. rer. nat.)

Verfasser:	Philipp Fuchs
Matrikel-Nummer:	0217430
Studienrichtung:	A 405 Mathematik
Betreuer:	O. Univ.-Prof. Dr. Walter Schachermayer

Wien, im Mai 2010

Acknowledgments

I was strongly supported by three persons. First of all I want to thank my adviser Walter Schachermayer, especially to make my visit at the summer school on Optimal Transport possible and for the academic contacts which made it possible to write this thesis.

Sincere thanks are given to Mathias Beiglböck for all the advise, discussions and his personal engagement. Especially I want to thank him (and the FWF) for the opportunity to be part of his research project (P21209), and last but not least for the numerous discussions beyond mathematics and academia.

Special thanks go to Max von Renesse who explained me patiently and in detail his research results, which were essential for my thesi. It expanded my point of view on his work tremendously and without his advise this thesis would not have been possible.

I am very grateful for the financial support by the university of Vienna that made my stay in Berlin possible.

Most of all, however, I want to thank my parents for the trust and all the years of support.

Schrödinger's Equation as Newton's Law of Motion

Philipp Fuchs

May 18, 2010

Contents

1 Introduction

The main goal of these notes is to present some recently developed techniques rooted in the field of optimal transport. These techniques allow to treat a class of PDE as gradient flows on the space of probability measures. First of all I should mention that the statements about the "Riemannian structure" of $\mathcal{P}(M)$ will be of a purely formal nature. This formal structure, however, will give us a more intuitive picture of the process studied. This process will be the dynamics of a quantum mechanical system, conventionally determined by the Schrödinger equation. I shall start with some results in optimal transport, important in the sequel, but also interesting in its own right. We will investigate the metric structure of $\mathcal{P}(M)$ and characterize the transport map for specific cost functions. Moreover, we will study the concept of displacement interpolation, a concept which points already towards the geometry of $\mathcal{P}(M)$. After that, a link to fluid mechanics will be provided, and we will derive the Benamou-Brenier formula. This formula is one of the main ingredients for our upcoming point of view. The purpose of the fist part is to derive this formula, and other prerequisites. Anyone just interested in the new techniques may skip this part and start right at Section 8. There a short summery is given, and I hope this will be sufficient to convey the main ideas. In the second part we will furnish $\mathcal{P}(M)$ with a formal Riemannian structure. We will study fundamental concepts of differential geometry (Gradient, Levi-Civita-connection...), and give three examples of interesting functionals and the flow induced by them. Then we will study the flow which is linked to quantum mechanics. The final statement will be, that the Schrödinger equation can be written as Newton's law of motion. As these notes are also the result of my personal endeavor to gain ground in a current field of interest, not all of the results and comments are of vital interest for the audience. I will try to give a remark at the beginning of every section, pointing the attention to the important results.

2 Optimal Transport

The main interest in Optimal Transport is, of course, to move mass in an optimal way. This means, given a cost function, to transport the mass in a way such that the required afford (cost) is minimal. One of the motivating examples mostly used is the sandpile, where we want to move a sandpile of volume, let us say 1, into a hole with the same volume. This question arises naturally and was of course considered before the year 1781. In this year, however, Gaspard Monge, a French mathematician, stated the problem in mathematical terms which is nowadays simply called the Monge-Problem. Given two probability spaces (X, μ) and (Y, ν), and a cost function $c : X \times Y \longrightarrow \mathbb{R}$ we are looking for a measurable map $T : X \longrightarrow Y$ such that

$T_{\#}\mu = \nu$ and

$$I[T] := \int_X c(x, T(x))\, d\mu(x)$$

is minimal. One of the problems in this case is, that, in general, such a map T does not exist. (For example if μ is a Dirac measure and ν is not, i.e. T as a map maps a point in X to exactly one point in Y, T can map a Dirac measure just onto another Dirac measure). In 1941 Leonid Kantorovich gave a more modern measure theoretical description which is a relaxed version of the Monge-Problem, i.e. if there exist a solution for the original problem there is also one in the Kantorovich framework, and they coincide. In fact the two problems are equal except that in the new one split of mass is allowed so we have no more problems with Dirac measures. The following is called the Kantorovich-Problem: Given two probability spaces (X, μ) and (Y, ν) and a cost function $c : X \times Y \longrightarrow \mathbb{R}$ we are looking for a probability measure π on the product space $X \times Y$, such that

$$\int_Y d\pi(x, y) = d\mu(x), \qquad \int_X d\pi(x, y) = d\nu(y),$$

or more precisely

$$\pi[A \times Y] = \nu[A], \qquad \pi[X \times B] = \nu[B] \tag{1}$$

for all measurable subsets A of X and B of Y. (Such a measure π is said to have marginals μ and ν.) And such that

$$I[\pi] := \int_{X \times Y} c(x, y)\, d\pi(x, y)$$

is minimal among all measures satisfying (1).

Remark 2.1. We will call a map T which satisfies $T_{\#}\mu = \nu$ a transport map, while we will call a $\pi \in \prod(\mu, \nu)$ a transference plan.

First I will give a few definitions and results.

Definition 2.2. (Semi-continuity) Let X be a topological space. A function
$f : X \longrightarrow \mathbb{R} \cup \{\pm\infty\}$ is called *lower semi- continuous* at x_0 if for all $\varepsilon > 0$ there exists a neighborhood U of x_0 such that $f(x) \geqslant f(x_0) - \varepsilon$. This is equivalent to

$$\liminf_{x \to x_0} f(x) \geq f(x_0).$$

The condition to be *upper semi-continuous* in x_0 is defined analogously, i.e.

$$\limsup_{x \to x_0} f(x) \leq f(x_0).$$

Definition 2.3. (Polish Space) A topological space is called a *Polish space* if it is separable and completely metrizable.

Definition 2.4. (Weak Convergence, [AGS08, 5.1]) A sequence of probability measures $(\mu_n) \subset \mathcal{P}(X)$ is *weakly* (or *narrowly*) convergent to $\mu \in \mathcal{P}(X)$ as $n \to \infty$ if

$$\lim_{n\to\infty} \int f(x)\, d\mu_n(x) = \int f(x)\, d\mu(x)$$

for all $f \in C_b^0(X)$, the set of continuous and bounded real-valued functions on X.

Theorem 1. ***(Prokhorov,[Bil99, Theorem 6.1, 6.2])*** *If a set $A \subset \mathcal{P}(X)$ is tight, i.e. for all $\varepsilon > 0$ there is a compact subset $K_\varepsilon \subset X$ such that $\mu(X \setminus K_\varepsilon) \leq \varepsilon$ for all $\mu \in A$, then A is relatively compact in $\mathcal{P}(X)$.*

Theorem 2. ***(Existence of Optimal Transference Plans, [Vil09, Theorem 4.1])*** *Let (X, μ) and (Y, ν) be two Polish probability spaces; let $a : X \longrightarrow \mathbb{R} \cup \{+\infty\}$ and $b : Y \longrightarrow \mathbb{R} \cup \{+\infty\}$ be two upper semi-continuous functions such that $a \in L^1(\mu)$, $b \in L^1(\nu)$. Let $c : X \times Y \longrightarrow \mathbb{R} \cup \{+\infty\}$ be a lower semi-continuous cost function, such that $c(x, y) \geq a(x) + b(y)$ for all x,y. Then there is a $\pi \in \prod(\mu, \nu)$ which minimizes the total cost*

$$\int_{X\times Y} c(x, y)\, d\pi(x, y)$$

among all possible $\pi \in \prod(\mu, \nu)$.

Sketch of the proof: First we show, that the cost functional $\int_{X\times Y} c\, d\pi$ is lower semi-continuous (l.s.c.) with respect to the weak convergence on $\mathcal{P}$(X), i.e.

$$\int_{X\times Y} c\, d\pi \leqslant \liminf_{n\to\infty} \int_{X\times Y} c\, d\pi_k \quad as\ \pi_k \to \pi.$$

This is a consequence of the fact, that a l.s.c. function can be written as the point wise supremum of a nondecreasing family of continuous real-valued functions. Next one shows that $\prod(\mu, \nu)$ is tight as $\{\mu\}$ is tight in $\mathcal{P}$(X), and $\{\nu\}$ is tight in $\mathcal{P}$(Y). By (1) we get that $\prod(\mu, \nu)$ is closed, and therefore $\prod(\mu, \nu)$ in fact is compact. Let now $(\pi_k)_k \in \mathbb{N}$ be a sequence such that $\int c\, d\pi_k$ converges to the infimum transport cost. Choosing a subsequence converging to some $\pi \in \prod(\mu, \nu)$ we see that

$$\int_{X\times Y} c\, d\pi \leqslant \liminf_{n\to\infty} \int_{X\times Y} c\, d\pi_k.$$

Thus π is a minimizer. The full proof can be found in [Vil09, 4.1].

$\square$

Remark 2.5. Probably the Monge-problem is what comes into ones mind at first thinking about an optimal transport strategy. However, this case would be much harder to handle. As shown above, thanks to Prokhorov's theorem, the existence of a minimizer in the Kantorovich-sense is quiet easy to prove. Of course at Monge's time no such fancy measure theory was available.

3 $\mathcal{P}(M)$ as a Metric Space

Now as we know about the existence of an optimal transference plan, we can approach the next step. That will be to study the transport problem where the cost function comes from a distance. We will see, that the cost functional fulfills the properties of a metric, and that it metrizes the weak topology of $\mathcal{P}$(X). There are other functionals metrizing the weak topology, but we will see that in the special case where $c(x,y) = \|x-y\|^2$ the metric enjoys some nice properties essential for our point of view.

Definition 3.1. (Wasserstein distances [Vil09, Definition 6.1]) Let (X,d) be a Polish metric space, and let $p \in [1,\infty)$. For any two probability measures μ, ν on X, the Wasserstein distance of order p between μ and ν is defined by the formula

$$\mathcal{W}_p := \left(\inf_{\pi \in \prod(\mu,\nu)} \int_{X \times X} d(x,y)^p \, d\pi(x,y) \right)^{\frac{1}{p}}.$$

Definition 3.2. (Wasserstein space, [Vil09, Definition 6.4]). Let (X,d) be a Polish, and let $p \in [1,\infty)$. The *Wasserstein space* of order p is defined as

$$\mathcal{P}_p(X) := \left\{ \mu \in \mathcal{P}(X), \quad \int_X d(x_0,x)^p \, \mu(dx) \; < \; +\infty \right\},$$

where $x_0 \in X$ is arbitrary.

Lemma 3.3. ***(Gluing lemma, [Vil03, Lemma 7.6]).*** *Let μ_1, μ_2, μ_3 be three probability measures, supported in Polish spaces X_1, X_2, X_3 respectively, and let $\pi_{12} \in \prod(\mu_1,\mu_2), \pi_{23} \in \prod(\mu_2,\mu_3)$ be two transference plans. Then there exists a probability measure $\pi \in \mathcal{P}(X_1,X_2,X_3)$ with marginals π_{12} on $X_1 \times X_2$ and π_{23} on $X_2 \times X_3$.*

Sketch of the proof:

The proof is an application of the technique of disintegration of measures. This allows to write a probability measure π on $X \times Y$ as $\int_X (\delta_x \otimes \pi_x) \, d\mu(x)$, where $\pi_x \in \mathcal{P}(Y) \; \forall x \in X$. Now we disintegrate π_{12} and π_{23} with respect to μ_2

$$\pi_{12} = \int_{X_2} \pi_{12,2} \otimes \delta_{x_2} \, d\mu_2(x_2) \qquad \pi_{23} = \int_{X_2} \delta_{x_2} \otimes \pi_{23,2} \, d\mu_2(x_2)$$

and define π by

$$\pi = \int_{X_2} (\pi_{12,2} \otimes \delta_{x_2} \otimes \pi_{23,2})\, d\mu_2(x_2).$$

$\square$

Lemma 3.4. ***(Minkowski's inequality, [Wer00, Korollar I.1.7])*** *Let $p \in [1, \infty)$ and f,g $\in L^p(X)$, then*

$$\|f + g\|_p \leqslant \|f\|_p + \|g\|_p$$

Theorem 3. ***(W_p is a metric, [Vil03, Theorem 7.3]).*** *W_p as defined in definition 3.1 defines a metric on $\mathcal{P}_p(X)$ as defined in 3.2.*

Sketch of the proof: The symmetry and nonnegativity of W_p follows from the nonnegativity and symmetry of d. It is clear that $W_p(\mu, \mu) = 0$, as Id is an optimal map. Conversely let $W_p(\mu, \nu) = 0$, then there is an optimal transference plan $d\pi(x, y)$ and it is supported on the diagonal (y=x). Thus $\forall \varphi \in C_b(X)$

$$\int \varphi\, d\mu = \int \varphi(x)\, d\pi_(x, y) = \int \varphi(y)\, d\pi_(x, y) = \int \varphi\, d\nu.$$

So $\mu = \nu$. It remains to show the triangular inequality. Let π be as in the gluing lemma, and π_{13} its marginal on $X \times Y$. This $\pi_{13} \in \prod(\mu_1, \mu_3)$. With this π in hand the proof is a straight forward chain of (in)equalities. *Optimal property of W_p, marginal property of π, triangular inequality of d and Minkowski's inequality for L^p.*

$\square$

The next result is interesting in its own right and will be a powerful tool.

Theorem 4. ***(Kantorovich duality, [Vil03, Theorem 1.3]).*** *Let X and Y be Polish spaces, let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, and let c: $X \times Y \longrightarrow \mathbb{R}_+ \cup \{+\infty\}$ be a lower semi-continuous cost function. Whenever $\pi \in \mathcal{P}(X \times Y)$ and $(\varphi, \psi) \in L^1(d\mu) \times L^1(d\nu)$, define*

$$I[\pi] = \int_{X \times Y} c(x, y)\, d\pi(x, y), \qquad J(\varphi, \psi) = \int_X \varphi\, d\mu + \int_Y \psi d\nu.$$

Define Φ_c to be the set of all measurable functions $(\varphi, \psi) \in L^1(d\mu) \times L^1(d\nu)$ satisfying

$$\varphi(x) + \psi(y) \leqslant c(x, y)$$

for $d\mu$-almost all $x \in X$, $d\nu$-almost all $y \in Y$.

Then

$$\inf_{\prod(\mu,\nu)} = \sup_{\Phi_c} J(\varphi,\psi) \tag{2}$$

and the infimum in the last equation is attained. One can also restrict the definition of Φ_c to those functions which are bounded and continuous, without changing the supremum.

About the proof (I just give a motivation how to proof the theorem, not a sketch of a proper proof as I did before.): One way to proof duality is to apply a so called minimax principle, i.e. to exchange *inf sup* by *sup inf.* To do this we first have to translate our linear constrained problem to a sup problem. Therefore we write:

$$\inf_{\pi\in\prod(\mu,\nu)} I[\pi] = \inf_{\pi\in M_+(X\times Y)} \left(I[\pi] + \begin{cases} 0 \text{ if } \pi \in \prod(\mu,\nu) \\ \infty \text{ else} \end{cases} \right)$$

Where $M_+(X \times Y)$ denotes the set of all nonnegative Borel measures This additional part we can write as:

$$\begin{Bmatrix} 0 \text{ if } \pi \in \prod(\mu,\nu) \\ \infty \text{ else} \end{Bmatrix} = \sup_{(\varphi,\psi)} \left[\int \varphi\, d\mu + \int \psi\, d\nu - \int [\varphi(x)+\psi(y)] d\pi(x,y) \right],$$

Now we can write the $\inf_{\pi\in\prod(\mu,\nu)}$ as:

$$\inf_{\pi\in M_+(X\times Y)} \sup_{(\varphi,\psi)} \left\{ \int_{(X\times\mathcal{Y}} c(x,y)\, d\pi(x,y) + \int_X \varphi\, d\mu + \int_Y \psi\, d\nu - \right.$$
$$\left. - \int_{(X\times\mathcal{Y})} [\varphi(x)+\psi(y)]\, d\pi(x,y) \right\}$$

Let us assume that a minimax principle can be applied, so that we get

$$\sup_{\varphi,\psi} \inf_{\pi\in M_+(X\times Y)} \left\{ \int_{(X\times\mathcal{Y}} c(x,y)\, d\pi(x,y) + \int_X \varphi\, d\mu + \int_Y \psi\, d\nu - \right.$$
$$\left. - \int_{(X\times\mathcal{Y})} [\varphi(x)+\psi(y)]\, d\pi(x,y) \right\}$$

This, however, is equal to:

$$\sup\,(\varphi,\psi) \left\{ \int_X \varphi\, d\mu + \int_Y \psi\, d\nu - \right.$$
$$\left. - \sup_{\pi\in M_+(X\times Y)} \int_{(X\times Y)} [\varphi(x)+\psi(y) \right\}.$$

What is now the value of this last sup (inside the curly brackets)? If $\varphi(x)+\psi(y)-c(x,y)$ is positive somewhere, we can define π such that the value of the integral becomes arbitrarily large (chose $\pi = c\delta_{(x,y)}$ and let $c \to \infty$). If $\varphi(x)+\psi(y)-c(x,y)$ is nonpositive, the sup is attained for $\pi = 0$. Hence:

$$\sup_{\pi \in M_+(X\times Y)} \int_{(X\times Y)} [\varphi(x)+\psi(y)-c(x,y)]\, d\pi(x,y) = \begin{cases} 0 \text{ if } (\varphi,\psi) \in \Phi_c \\ \infty \text{ else} \end{cases}$$

And this finally concludes the proof as we see:

$$(3) = \sup_{(\varphi,\psi)\in\Phi_c} J(\varphi,\psi)$$

□

Remark 3.5. It is easy to see, that

$$\sup_{\Phi_c \cap C_b} J(\varphi,\psi) \le \sup_{\Phi_c \cap L^1} J(\varphi,\psi) \le \inf_{\prod(\mu,\nu)} I[\pi].$$

This is simply a consequence of

$$\varphi(x)+\psi(y) \le c(x,y)$$

and the marginal property of π.

Remark 3.6. In the new book of Villani ([Vil03]) the proof is performed using the concept of cyclical monotonicity and c-convex functions (c-concave functions, respectively). We shall be concerned with this notions below. The optimal transport problem and the proof of Kantorovich duality are also motivated and explained by a nice heuristic. So for better understanding this source might be suggested. For the present notes however, I decided to present this one to stay consistent with the proofs coming up in the sequel.

Remark 3.7. It follows, that for bounded cost functions the set Φ_c in the right-hand side of (9) can be restricted to the set φ^{cc}, φ^c, where φ is bounded and:

$$\varphi^c(y) = \inf_{x\in X}[c(x,y)-\varphi(x)], \qquad \varphi^{cc} = \inf_{y\in Y}[c(x,y)-\varphi^c(y)]. \tag{3}$$

The pair (φ^{cc},φ^c) is called a pair of conjugate c-concave functions.

Theorem 5. ***(Kantorovich-Rubinstein theorem, [Vil03, Theorem 1.14]).*** *Let X=Y be a Polish space and let d be a metric on X. Let $\mathcal{T}_d$ be the cost of optimal transportation for the cost $c(x,y) = d(x,y)$,*

$$\mathcal{T}_d = \inf_{\pi\in\prod(\mu,\nu)} \int_{X\times X} d(x,y)\, d\pi(x,y).$$

Let Lip(X) denote the space of all Lipschitz functions on X, and

$$\|\varphi\|_{Lip} \equiv \sup_{x \neq y} \frac{|\varphi(x) - \varphi(y)|}{d(x,y)}.$$

Then

$$\mathcal{T}(\mu,\nu) = \sup \left\{ \int_X \varphi \, d(\mu - \nu); \quad \varphi \in L^1(d|\mu - \nu|); \quad \|\varphi\|_{Lip} \leqslant 1 \right\}.$$

Moreover it does not change the value of the supremum above to impose the additional condition that φ be bounded.

Sketch of the proof: The proof is a first application of the duality-theorem. First, however, one show that one can assume that d is bounded. Then we only have to check, that:

$$\sup_{(\varphi,\psi)} \in \Phi_d \; J(\varphi,\psi) = \left\{ \int_X \varphi \, d(\mu - \nu); \quad \|\varphi\|_{Lip} \leqslant 1 \right\}$$

Recall that our cost function $c(x,y)$ is given by $d(x,y)$, this explains the notion Φ_d. The last equality can be deduced from Remark 3.5 above, i.e

$$\sup_{(\varphi,\psi)\in\Phi_d} J(\varphi,\psi) = \sup_{\varphi \in L^1(d\mu)} J(\varphi^{dd}, \varphi^d)$$

and well known properties of Lipschitz functions.

□

Now we are prepared to approach the final result of this chapter. We will show, that W_p corresponds to the weak topology of $\mathcal{P}_P(X)$.

Theorem 6. ***(Metrization of Weak Convergence, [Vil03, Theorem 7.12])*** *Let $p \in [1,\infty)$, let $(\mu_k)_{k\in\mathbb{N}}$ be a sequence of probability measures in $\mathcal{P}_p(X)$, and let $\mu \in \mathcal{P}(X)$. Then, the following are equivalent:*

(i) $W_p(\mu_k,\mu) \xrightarrow[k\to\infty]{} 0.$

(ii) $\mu_k \xrightarrow[k\to\infty]{} \mu$ *in the weak sense, and $(\mu_k)_k \in \mathbb{N}$ satisfies the following condition:*

$$\lim_{R\to\infty} \limsup_{k\to\infty} \int_{d(x_0,x)\geqslant R} d(x_0,x)^p \, d\mu_k(x) = 0$$

for some $x_0 \in X$.

(iii) $\mu_k \xrightarrow[k\to\infty]{} \mu$ *in weak sense, and there is convergence of the moment of order p:*

$$\int d(x_0,x)^p\, d\mu_k(x) \xrightarrow[k\to\infty]{} \int d(x_0,x)^p\, d\mu(x).$$

for any $x_0 \in X$.

(iv) Whenever a continuous function φ *on* X *satisfies the growth condition* $|\varphi(x)| \leqslant C[\,1 + d(x_0,x)^p]$ *for some* $x_0 \in X$, $C \in \mathbb{R}$, *then*

$$\int \varphi\, d\mu_k \xrightarrow[k\to\infty]{} \int \varphi\, d\mu.$$

Sketch of the proof: The interesting, and therefore of course difficult, part in the proof is the equivalence of (i) and (iii). The other equivalences can be deduced rather quickly. First of all, the weak convergence implies the inequality

$$\int d(x_0,x)^p\, d\mu(x) \le \liminf_{k\to\infty} \int d(x_0,x)^p\, d\mu_k(x),$$

thus, once the weak convergence granted, it suffices to show, that

$$\limsup_{k\to\infty} \int d(x_0,x)^p\, d\mu_k(x) \le \int d(x_0,x)^p\, d\mu(x).$$

The next step is to show that convergence in the W_p sense implies this inequality. Combining the regular triangular inequality and

$$(a+b)^p \le (1+\varepsilon)a^p + C_\varepsilon d(x,y)^p$$

and integrating over π_k (i.e. an optimal transference plan between μ_k and μ, we get:

$$\int d(x_0,x)^p\, d\mu_k(x) \le (1+\varepsilon)\int d(x_0,x)^p\, d\mu(y) + C_\varepsilon \int d(x,y)^p\, d\pi_k(x,y)$$

in the integral in the right hand side we integrate over x, so we get by (1) the integral with resp. to μ_k, and μ for the second integral where we integrate over y. As the second summand on the left hand side is just the Wasserstein distance between μ_k and μ, and we assumed convergence in that sense, we find that

$$\limsup_{k\to\infty} \int d(x_0,x)^p\, d\mu_k(x) \le (1+\varepsilon)\int d(x_0,x)^p\, d\mu(x).$$

If we now let $\varepsilon \to 0$ we obtain the desired inequality. In short: If W_p implies the weak convergence it implies automatically the convergence of the p-th moments. Now it remains to show that W_p really implies the weak convergence and that (iii) implies (i). To finish the proof we switch to the distance $\tilde{d} := \inf(d, 1)$ and investigate $\tilde{W}_p$. It turns out that convergence in the $\tilde{W}_p$ sense implies the convergence in the W_p sense. As $W_p \geq \tilde{W}_p$ this concludes the proof. Let us now assume that d is bounded, e.g. $d \leq 1$. As now all the distances are equivalent we concentrate on the special case where $p = 1$, and we can apply the Kantorovich-Rubinstein theorem. Convergence in the W_p sense now reduces to

$$\sup_{\|\varphi\|_{Lip} \leqslant 1} \int \varphi \, d(\mu_k - \mu) \xrightarrow[k\to\infty]{} 0. \tag{4}$$

Assume now that $W_1(\mu_k, \mu) \to 0$. We have to show that

$$\int \varphi \, d\mu_k \xrightarrow[k\to\infty]{} \int \varphi \, d\mu \tag{5}$$

for all $\varphi \in C_b(X)$. Now (4) implies that (5) is true for 1-Lipschitz functions. Replacing φ by $\frac{\varphi}{\|\varphi\|}$ we find that the statement is true for general Lipschitz functions. We here recall, that bounded continuous functions on metric spaces can be approximated from above and below by Lipschitz functions. (See for example [Mic01]).

Finally we have to establish that $(i) \Rightarrow (iii)$. It suffices to show that

$$\sup_{\varphi \in Lip_{1,x_0}} \int \varphi \, d(\mu_k - \mu) \xrightarrow[k\to\infty]{} 0.$$

Where Lip_{1,x_0} is the set of Lipschitz functions with Lipschitz constant at most 1, and $\varphi(x_0) = 0$. Prokhorov's theorem provides us a convenient family of compact sets K_n and Ascoli's theorem (a proof can be found in [KN76])with a convenient sequence of functions φ_k, such that it only remains to show that

$$\int \varphi_k \, d(\mu_k - \mu) \xrightarrow[k\to\infty]{} 0.$$

Using the properties of our family K_n and the assumed weak convergence of μ_k we get the desired result.

□

Remark 3.8. Now we know, that W_P metrizes $\mathcal{P}_p(X)$, the set of probability measures with finite p-th moment. Moreover, if d is bounded, then W_p metrizes the weak topology of $\mathcal{P}$(X). As we always can replace d by a equivalent (bounded) distance, it follows that $\mathcal{P}$(X) itself is a metric space.

Remark 3.9. To get a better insight how W_p metrizes the weak topology one may take a look into chapter 7.3 of [Vil03] where the proof of the theorem is given for $\mathbb{R}$ in terms of cumulative distribution functions. In [Vil09, 6.18] it can is shown that $\mathcal{P}_p(X)$ itself is a Polish space if X is one.

Maybe this is a good point for a break, and to sum up what we have achieved so far. The idea, that the minimal afford to transport one measure onto another one can be seen as a distance between them, seems quiet obvious. However, to see that this is consistent with the weak topology, and to justify it in full mathematical rigor requires some work. The importent theorems are, of course, 4 and 5. Duality will used subsequently. The purpose of it (i.e. to have a formulation in terms of functions) for optimal transport should not be underestimated. We tried to introduce to the reader the central ideas of the proofs in the sketches to provide a guideline ignoring too technical details.

4 Transport Maps, Existence and Uniqueness

The goal of this section is to give a theorem which links transference plans to transport maps. We only elaborate on the case of the quadratic cost function; it is the one important for us in the sequel. First we will show that the set of functions admissible in the dual problem can be restricted to a smaller set. To do so we will take advantage of the specific structure of the quadratic cost. Throughout this section we set $X = Y = \mathbb{R}^n$ if nothing else is said. Our cost function will not be $c = \|x - y\|^2$, but $c = \frac{\|x-y\|^2}{2}$ (note that this does not effect the MKP, but is convenient for our calculations). To be admissible in the dual problem means

$$\varphi(x) + \psi(y) \leq \frac{\|x - y\|^2}{2}.$$

Expanding the right-hand side we get

$$\varphi(x) + \psi(y) \leq \frac{\|x\|^2}{2} + \frac{\|y\|^2}{2} - 2\frac{x \cdot y}{2},$$

where $\cdot$ denotes the inner product. After rearranging terms we find

$$x \cdot y \leq \left[\frac{\| x \|^2}{2} - \varphi(x)\right] + \left[\frac{\| y \|^2}{2} - \psi(y)\right]. \tag{6}$$

Anyone familiar with convex analysis might already recognize the analogy to the theory of convex conjugate functions. We set

$$\widetilde{\varphi}(x) := \frac{\| x \|^2}{2} - \varphi(x) \qquad \widetilde{\psi}(y) := \frac{\| y \|^2}{2} - \psi(y). \tag{7}$$

For notational convenience, moreover we define

$$M_2 := \int \frac{\| x \|^2}{2}\, d\mu(x) + \int \frac{\| y \|^2}{2}\, d\nu(y) \tag{8}$$

and assume that, μ, ν are of finite second order, i.e. $M_2 < +\infty$. Recall that there is no duality gap, i.e.

$$\inf_{\pi \in \prod(\mu,\nu)} I[\pi] = \sup_{(\varphi,\psi)\in\Phi_c} J(\varphi,\psi). \tag{9}$$

Next we rephrase our optimality condition as

$$\inf_{\pi \in \prod(\mu,\nu)} I[\pi] = \inf_{\pi \in \prod(\mu,\nu)} \int \frac{\|x\|^2}{2} + \frac{\|y\|^2}{2} - x \cdot y \, d\pi = M_2 - \inf_{\pi \in \prod(\mu,\nu)} \int x \cdot y \, d\pi =$$

$$= M_2 - \sup_{\pi \in \prod(\mu,\nu)} \int x \cdot y \, d\pi \tag{10}$$

we now define $\tilde{\Phi}_c$ as the set that consits of all pairs $(\tilde{\varphi}, \tilde{\psi}) \in L^1(d\mu) \times L^1(d\nu)$, s. t. for all x, y

$$x \cdot y \leq \tilde{\varphi}(x) + \tilde{\psi}(y). \tag{11}$$

We get that

$$\sup_{\Phi_c} J = \sup_{\Phi_c} \int M_2 - \tilde{\varphi}(x) - \tilde{\psi}(y) =$$

$$= M_2 - \inf\{J(\tilde{\varphi}, \tilde{\psi}); \quad (\tilde{\varphi}, \tilde{\psi}) \in \widetilde{\Phi}\}, \tag{12}$$

After this calculations we see that we get the minimal value of our transport problem, if we find the inf in (12). Now, via the double convexification trick, we will shrink the set $\widetilde{\Phi}$ to a much smaller one. Note that due to (11)

$$\tilde{\psi}(y) \geq \sup_x [x \cdot y - \tilde{\varphi}(x)] =: \tilde{\varphi}^*(y). \tag{13}$$

With (13) in hand we arrive at

$$J(\tilde{\varphi}, \tilde{\psi}) \geq J(\tilde{\varphi}, \tilde{\varphi}^*). \tag{14}$$

Next, for all $x \in X$,

$$\tilde{\varphi}(x) \geq \sup_y [x \cdot y - \tilde{\varphi}^*(y)] =: \tilde{\varphi}^{**}$$

and therefore

$$.J(\tilde{\varphi}, \tilde{\varphi}^*) \geq J(\tilde{\varphi}^{**}, \tilde{\varphi}^*) \tag{15}$$

Combining (14) and (15).

$$\inf_{(\tilde{\varphi},\tilde{\psi})\in\widetilde{\Psi}} J(\tilde{\varphi}, \tilde{\psi}) \geq \inf_{\tilde{\varphi}\in L^1(d\mu)} J(\tilde{\varphi}^{**}, \tilde{\varphi}^*) \tag{16}$$

Admitting that $(\tilde{\varphi}^{**}, \tilde{\varphi}^*) \in L^1(d\mu) \times L^1(d\nu)$ one finds that the inf of J over $\widetilde{\Phi}$ is the same as the inf over the subset consisting of the pairs $(\tilde{\varphi}^{**}, \tilde{\varphi}^*)$. We here will recall some facts about convex analysis, but before we make the important

Remark 4.1. If we assume that (11) holds only true for almost all x, y we can modify φ and ψ in the following way. Let N_x and N_y denote the sets where the equation does not hold true. Set φ and ψ to be $+\infty$ on this sets. Now the equation holds true for all x and y and as $N_x \times N_y$ is a π zero set the value of $J(\varphi, \psi)$ is unchanged, and (φ, ψ) still belong to $\tilde{\Phi}$.

Remark 4.2. The equation (6) is essential for us, as we want to apply results of convex analysis. With this tools we will finally be able to prove the existience of a transport map.
From now on, however, we will drop the $\sim$ symbol on $(\tilde{\varphi}, \tilde{\psi})$ in (7).

Definition 4.3. (Convex conjugate functions) For any proper (not identically $+\infty$) function $\varphi : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ one can define it's convex conjugate function, also called Legendre transform, by

$$\varphi^* = \sup_{x\in\mathbb{R}^n} (x \cdot y - \varphi(x)) \tag{17}$$

Remark 4.4. Note that only because of the special structure of the quadratic cost function we can take advantage of the Legendre transform. For a general cost function the concept has to be modified.

These pairs of convex lower semi-continuous functions give reason to recall some facts about convex analysis. An exhaustive source on the matter is [Roc97]. One may also take a look into [EG92]. We will just briefly quote some facts which are needed subsequently. We begin with the

Definition 4.5. (proper convex function) A proper convex function φ on $\mathbb{R}^n$ is a function $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, not identically $+\infty$, such that

$$\forall x, y \in \mathbb{R}^n, \forall t \in [0, 1], \qquad \varphi(tx + (1 - t)y) \leq t\varphi(x) + (1 - t)\varphi(y).$$

We proceed giving some facts: The gradient of a proper convex function is well-defined almost everywhere (Rademacher's theorem [EG92]). The graph lies above the tangent, just as one imagines in the one dimensional case. In the points where the function φ is not differentiable one can define the so called subdifferential ($\partial\varphi$). More precisely we have

$$y \in \partial\varphi \Leftrightarrow [\forall z \in \mathbb{R}^n, \quad \varphi(z) \geq \varphi(x) + \langle y, z - x\rangle]. \tag{18}$$

In one dimension, although the derivative in a point might not exist, the right and the left limit of the differential do exist. In that case, the subdifferential is the set of all values between these two limits. The next proposition provides more information about the subdifferential.

Proposition 4.6. ***(Characterization of subdifferential, [Vil03, Proposition 2.4])*** *Let φ be a proper lower semi-continuous convex function on $\mathbb{R}^n$. Then for all x,y $\in \mathbb{R}^n$,*

$$x \cdot y = \varphi(x) + \varphi^*(y) \Longleftrightarrow y \in \partial\varphi(x) \Longleftrightarrow x \in \partial\varphi^*(y).$$

Remark 4.7. Note that, if φ is differentiable, the Legndre transform is defined by $\varphi^*(x) = x \cdot \nabla\varphi(x) - \varphi(x)$. So the above proposition does not really come as a surprise.

The next proposition states that for lower semi-continuous functions the Legendre transform is its own inverse.

Proposition 4.8. ***(Legendre duality, [Vil03, Proposition 2.5])*** *Let $\varphi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be a proper function. Then the following three properties are equivalent:*

(i) *φ is convex and lower semi-continuous.*

(ii) *$\varphi = \psi^*$ for some proper function ψ.*

(iii) *$\varphi^{**} = \varphi$.*

The proof (just a few lines) can be found in the references. Of course, much more could be said about convex analysis. The above mentioned, however, will hopefully cover everything needed subsequently. Before we can tackle the desired theorem of this section we need one more result. It states that there exists a pair of optimal convex conjugate functions. Besides being a crucial tool for us, it is of interest on its own. To proof it we need another result which is not (of interest on its own) but a technicality. The double convexification lemma.

Lemma 4.9. ***(Double convexification lemma, [Vil03, Lemma 2.10])*** *Let μ, ν be probability measures respectively supported in subsets X and Y of $\mathbb{R}^n$, satisfying*

$$M_2 \equiv \int_X \frac{|x|^2}{2}\, d\mu(s) + \int_Y \frac{|y|^2}{2}\, d\nu(y) < +\infty.$$

Whenever φ, ψ are measurable functions with values in $\mathbb{R} \cup \{+\infty\}$, introduce

$$\varphi^*(y) = \sup_{x \in X} [x \cdot y - \varphi(x)], \tag{19}$$

$$\psi^*(x) = \sup_{y \in Y} [x \cdot y - \psi(y)]. \tag{20}$$

Let $\tilde{\Phi}$ defined by (11), and let $(\varphi_k, \psi_k)_{k \in \mathbb{N}}$ be a minimizing sequence for J on $\tilde{\Phi}$. Then, (i) One can modify $(\varphi_k, \psi_k)_{k \in \mathbb{N}}$ on zero-measure sets (with respect to μ, ν) in such a way that inequality (11) holds true for all (x,y) $\in \mathbb{R}^n \times \mathbb{R}^n$, without changing the values J(φ_k, ψ_k). (ii) There exists a sequence of real numbers $(a_k)_{k \in \mathbb{N}}$ such that

$$(\overline{\varphi}_k, \overline{\psi}_k) = (\varphi_k^{**} - a_k, \varphi_k^* + a_k) \tag{21}$$

is still a minimizing sequence for J on $\tilde{\Phi}$, and satisfies the lower bounds

$$\forall x \in X, \forall y \in Y, \qquad \overline{\varphi}_k(x) \geqslant -\frac{|x|^2}{2},\ \overline{\psi}_k(x) \geqslant -\frac{|y|^2}{2}, \tag{22}$$

together with the "upper bounds"

$$\liminf_{k \to \infty} \inf_{x \in X} \left(\overline{\varphi}_k(x) + \frac{|x|^2}{2} \right) \leqslant \inf_{\Phi} J + M_2 \tag{23}$$

$$\liminf_{k \to \infty} \inf_{y \in Y} \left(\overline{\psi}_k(y) + \frac{|y|^2}{2} \right) \leqslant \inf_{\Phi} J + M_2 \tag{24}$$

*(iii) In particular, with the choice X=Y=$\mathbb{R}^n$, the * operation coincides with the usual Legendre transform, and*

$$\inf_{\Phi} J = \inf_{\varphi \in L^1(d\mu)} J(\varphi^{**}, \varphi^*).$$

So the infimum of J on $\tilde{\Phi}$ does not change upon restriction J to the narrow set of $\tilde{\Phi}$ made of pairs of conjugate proper convex functions.

Sketch of the proof: Note first that we already assumed that (11) holds true for all and not just for almost all x and y (Remark 4.1). (iii) follows directly from (ii), therefore it only remains to show (ii). First because φ_k and ψ_k are proper (i.e. $\not\equiv +\infty$), we can give lower bounds on them. This allows us to give (for every k) a finite a_k. We now choose our modified sequence to be as in (21). Note now, that

$$\forall a \in \mathbb{R}, \quad J(\varphi + a, \psi - a) = J(\varphi, \psi) \tag{25}$$

and $\overline{\varphi}_k = (\overline{\varphi}_k)^*$. To show that this sequence fulfills (22) is merely a computation. Because of (25) and properties of convex conjugate functions we obtain

$$J(\overline{\varphi}_k, \overline{\psi}_k) = J(\varphi_k^{**}, \varphi_k^*) \leq J(\varphi_k, \psi_k) < +\infty.$$

Admitting the integrability of $\overline{\varphi}_k$ and $\overline{\psi}_k$ we see that our modified sequence is minimizing as well . The last two conditions ((23) and (24)) are checked by computation.

□

Now the promised existence result.

Theorem 7. ***(Existence of an optimal pair of convex conjugate functions, [Vil03, Theorem 2.9]).*** *Let μ and ν be two probability measures on $\mathbb{R}^n$, with finite second order moments. Let $\widetilde{\Phi}$ be defined as in* (11). *Then, there exists a pair (φ, φ^*) of lower semi-continuous proper conjugate convex functions on $\mathbb{R}^n$, such that*

$$\inf_{\widetilde{\Phi}} J = J(\varphi, \varphi^*)$$

I will give a sketch of the proof in the case where the probability measures μ and ν are supported in compact subsets of $\mathbb{R}^n$. In this case almost all the work was done in the previous lemma. The result, indeed, holds true in a much more general setting.

Sketch of the proof: Let now X, Y $\subset \mathbb{R}^n$ compact, and μ, ν be supported upon them. It is *easy* to show that the functions $\overline{\varphi}_k$ and $\overline{\psi}_k$ of our modified minimizing sequence are uniformly Lipschitz and uniformly bounded. According to Ascoli's theorem there exist subsequences of $\overline{\varphi}_k$ and $\overline{\psi}_k$ converging uniformly in $C_b(X), C_b(Y)$ to continuous limits $\overline{\varphi}$, $\overline{\psi}$. This pair is still optimal and we extend them outside X,Y by $+\infty$ and double convexify them. This concludes the proof.

□

We are now prepared for the desired statement of this section.

Theorem 8. ***(Optimal transportation theorem for quadratic cost, [Vil03, Theorem 2.12])*** *Let μ, ν be probability measures on $\mathbb{R}^n$, with finite second moments. We consider the Monge-Kantorovich transportation problem associated with the quadratic cost function c(x,y)=$|x-y|^2$. Then,*

(i) $\pi \in \prod(\mu,\nu)$ is optimal if and only if there exists a convex lower semi-continuous function φ such that

$$Supp(\pi) \subset Graph(\partial\varphi), \tag{26}$$

or equivalently:

$$for\, d\pi - almost\ all\ (x,y), \quad y \in (\partial\varphi(x)). \tag{27}$$

Moreover, in that case, the pair (φ, φ^*) has to be a minimizer in the problem

$$\inf\left\{\int_{\mathbb{R}^n} \varphi\, d\mu + \int_{\mathbb{R}^n} \psi\, d\nu; \quad \forall (x,y), \quad x \cdot y \leqslant \varphi(x) + \psi(y)\right\}.$$

(ii) If μ dose not give mass to small sets (see Remark 4.10 below), then there is a unique optimal π, which is

$$d\pi(x,y) = d\mu(x)\delta[y = \nabla\varphi(x)], \tag{28}$$

or equivalently,

$$\pi = (Id \times \nabla\varphi)_{\#}\mu, \tag{29}$$

where $\nabla\varphi$ is the unique (i.e. uniquely determined $d\mu$ almost everywhere) gradient of a convex function which pushes μ forward toν: $\nabla\varphi_{\#}\mu = \nu$. Moreover,

$$Supp(\nu) = \overline{\nabla\varphi(Supp(\nu))}.$$

(iii) As a corollary, under the assumptions of (ii), $\nabla\varphi$ is the unique solution to the Monge transportation problem:

$$\int_{\mathbb{R}^n} |x - \nabla\varphi(x)|^2\, d\mu(x) = \inf_{T_{\#}\mu=\nu} \int_{\mathbb{R}^n} |x - T(x)|^2\, d\mu(x),$$

or equivalently,

$$\int_{\mathbb{R}^n} x \cdot \nabla\varphi\, d\mu(x) = \sup_{T_{\#}\mu=\nu} \int_{\mathbb{R}^n} |x \cdot T(x)\, d\mu(x).$$

(iv) Finally, if ν also does not give mass to small sets, then, for $d\mu$-almost all x and $d\nu$-almost all y,

$$\nabla\varphi^* \circ \nabla\varphi(x) = x, \qquad \nabla\varphi \circ \nabla\varphi^* = y,$$

and $\nabla\varphi^*$ is the ($d\nu$-almost everywhere) unique gradient of a convex function which pushes ν forward to ν, and also the solution of the Monge problem for transporting ν onto μ with a quadratic cost function.

Remark 4.10. In convex analysis a notion of small sets it to be of Hausdorff dimension at most $n-1$. If one does not feel convenient with this notion think of small sets as Lebesgue zero sets.

Sketch of the proof: The equivalences (26) $\Leftrightarrow$ (27) and (28) $\Leftrightarrow$ (29) are obvious. Recall now that due to the calculations at the beginning of this section we can reduce our minimizing problem to (10) - (11).

(i) We know that there exists an optimal transference plan π (theorem 2) and an optimal pair of convex conjugate functions (φ, φ^*) (theorem 7). Recalling the marginal property of π we see that

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} (x \cdot y)\, d\pi(x,y) = \int_{\mathbb{R}^n \times \mathbb{R}^n} [\varphi(x) + \varphi^*(y)]\, d\pi(x,y).$$

Rearranged this is

$$0 = \int_{\mathbb{R}^n \times \mathbb{R}^n} [\varphi(x) + \varphi^*(y) - x \cdot y]\, d\pi(x,y).$$

From (17) we now that

$$\forall x, y \in \mathbb{R}^n, \quad x \cdot y \leq \varphi(x) + \varphi^*(y).$$

Hence the integrand above is nonnegative and therefore has to vanish π-almost everywhere. Recalling proposition 4.6 this entails (27). If now conversely $\pi \in \prod(\mu, \nu)$ satisfies (27), then

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} (x \cdot y)\, d\pi = \int_{\mathbb{R}^n} \varphi\, d\mu + \int_{\mathbb{R}^n} \varphi^*\, d\nu.$$

And this means that both π and (φ, φ^*) are optimal.

(ii) Since μ does not give mass to small sets (think of small sets as Lebesgue zero sets) and $\varphi \in L^1(d\mu)$ and φ is a proper convex function we see that $\mu[Int(Dom(\varphi))] = 1$. Moreover, as the set of points in the interior of the domain where φ is not differentiable is a small set the set of differentiability points of X is of full μ-measure. This means that $\partial\varphi$ consists of just one point $\{\nabla\varphi(x)\}$ μ-a.e., thus $y = \nabla\varphi(x)$ π-a.e. Now we want to show the uniqueness of our optimal π. Where $\pi = (Id \times \nabla\varphi)_{\#}\mu$ for some convex φ s. t. $\nabla\varphi = \nu$. Assume that there is another convex function $\overline{\varphi}$ s. t. $\nabla\overline{\varphi}_{\#}\mu = \nu$. These functions induce an optimal pair for the dual problem via convex conjugation, i.e.

$$\begin{aligned}\int_{\mathbb{R}^n \times \mathbb{R}^n} [\overline{\varphi}(x) + \overline{\varphi}^*(y)]\, d\pi(x,y) &= \int_{\mathbb{R}^n \times \mathbb{R}^n} [\varphi(x) + \varphi^*(y)]\, d\pi(x,y) \\ &= \int_{\mathbb{R}^n \times \mathbb{R}^n} (x \cdot y)\, d\pi(x,y).\end{aligned}$$

As $\pi = (Id \times \nabla\varphi)_{\#}\mu$, after rearranging terms we see that

$$\int_{\mathbb{R}^n} [\overline{\varphi}(x) + \overline{\varphi}^*(x) - x \cdot \nabla\varphi(x)]\, d\mu(x) = 0.$$

With the same arguments as before we see that the integrand is non-negative and we end up with

$$\nabla\varphi(x) = \nabla\overline{\varphi}(x) \quad d\pi\text{-a.e.}$$

The equality

$$Supp(\nu) = \overline{\nabla\varphi(Supp(\nu))}$$

again is a consequence of the convexity of φ. There is nothing more to do now to gain (iii), and (iv) holds true for the same reasons as (i) and (ii) using that (φ, φ^*) is a pair of proper functions.

□

This theorem is of great use for us as we finally have deduced the existence of a unique transport map and information about its shape. It can be extended to the much more general case of a strictly convex cost function. Then however the transport map looks differently. We will not elaborate on this fact, as for us the quadratic cost is of main interest. Nevertheless, we want to generalize for Riemannian manifolds, the natural setting for our upcoming investigations.

Theorem 9. ***(McCann's theorem, [Vil03, Theorem 2.47]).*** *Let M be a connected, complete smooth Riemannian manifold, equipped with its standard volume measure dx. Let μ, ν be two probability measures on M with compact support, and let the cost function be c(x,y) $= d(x,y)^2$, where d is the geodesic distance on M. Further, assume that μ is absolutely continuous with respect to the volume measure on M. Then, The Monge-Kantorovich problem between μ and ν admits a unique optimal transference plan, and it has the form $d\pi(x,y) = d\mu(x)\delta[y = T(x)]$, or equivalently*

$$\pi = (Id \times T)_{\#}\mu),$$

where T is uniquely determined, μ- almost everywhere, by the requirements that $T_{\#}\mu = \nu$ and

$$T(x) = exp_x[-\nabla\varphi(x)]$$

for some $d^2/2$-concave function φ.

This is a slightly simplified version of McCann's original result. See [McC01].

5 The Time-Dependent Version

It has hardly to be motivated that a solution to our problem depending on time is desirable. It would give us an insight how the transport would have to be performed for practical purpose and provide a richer theory with a variety of applications. Recall that in the case of our interest we already have shown the existence of an optimal transport map, solving the Monge problem, i.e.

$$\inf \left\{ \int_X c(x, T(x))\, d\mu(x); \quad T_\# \mu = \nu \right\}. \tag{30}$$

Our approach to acquire information about the history of our transportation process is to investigate the trajectories. This means that we associate to each x (particle) a function $(T_t(x))_{0 \le t \le 1}$ (path). The cost of transporting a single particle along its trajectory is denoted by $C[T_t(x))]$. In this new formulation our problem is to find

$$\inf \left\{ \int_X C[(T_t(x))_{0 \le t \le^1}]\, d\mu(x); \quad T_0 = Id,\, T_{1\#} \mu = \nu \right\}. \tag{31}$$

What we want is that (30) and (31) predict the same transportation cost, transportation map respectively. A simple and natural condition therefore is

$$c(x, y) = \inf \left\{ C[(z_t)_{0 \le t \le 1}]; \quad z_0 = x,\ z_1 = y \right\}. \tag{32}$$

In many cases of interest the cost is of the form

$$C[(z_t)] = \int_0^1 c(\dot{z}_t)\, dt$$

An prominent example is, where the cost is the energy associated to the path (forget about the constant $\frac{1}{2}$)

$$C[(z_t)] = \int_0^1 \parallel \dot{z} \parallel^2 \ dt \quad in\, \mathbb{R}^n \quad \Rightarrow c(x, y) = \parallel x - y \parallel^2 .$$

This is a special case of the next proposition.

Proposition 5.1. ***(Extremal trajectories for convex costs are straight lines [Vil03, Proposition 5.2])*** *If c is a convex function on $\mathbb{R}^n$, then*

$$\inf \left\{ \int_0^1 c(\dot{z}_t)\, dt; \quad z_0 = x,\, z_1 = y \right\} = c(x - y).$$

or, $[0,T]$*-parametrized*

$$\inf\left\{\int_0^T c(\dot{z}_t)\,dt;\quad z_0 = x,\ z_T = y\right\} = Tc\left(\frac{y-x}{T}\right).$$

If, moreover, c is strictly convex, then the inf is achieved uniquely by

$$z_t = x + t(y-x) \qquad x + \frac{t}{T}(y-x) \text{ respectively.}$$

The proof is an essential consequence of Jensen's inequality.

Lemma 5.2. ***(Jensen's inequality [Els05, VI, Lemma 1.3])*** *Let* (X,μ) *be a probability space,* $I \subset \mathbb{R}$ *an interval,* $f : X \to I$ μ*-integrable and* $\varphi : I \to \mathbb{R}$ *convex. Then* $\int_X f\,d\mu \in I$*,* $\varphi \circ f$ *is quasi integrable, and*

$$\varphi\left(\int_X f\,d\mu\right) \leq \int_X \varphi \circ f\,d\mu.$$

By (32) we see that the trajectories for almost all x have to be optimal and, moreover, for convex functions these trajectories have to be straight lines for for displacement costs of the form $C[(z_t)] = \int_0^1 \parallel \dot{z} \parallel^2 dt$. These conditions motivate (and already proof) the following theorem.

Theorem 10. ***(Time-dependent optimal transportation theorem [Vil03, Theorem 5.5])*** *Consider the cost function* $c(x,y) = \parallel x - y \parallel$ *in* $\mathbb{R}^n$*. Let* μ, ν *be probability measures with finite second moments, and let* $C[(z_t)] = \int_0^1 c(\dot{z}_t)dt$*. Let* $\nabla\varphi$ *be as defined in Theorem 8. Then the solution to our time-dependent problem is given by:*

$$T_t(x) = x - t\nabla\varphi(x), \qquad 0 \leq t \leq 1.$$

There are still two things we are interested in. Is the result also true for Riemannian manifolds (one should think about the cut locus), and is the transport in between optimal? Recall that every T_t defines a measure via the push forward with respect to the reference measure $T_{t\#}\mu = \mu_t$. The question is, if the transport between μ and μ_t is also optimal. The next proposition will give a positive answer to both questions.

Theorem 11. ***(Intermediate time optimality theorem)[Vil03, Theorem 5.6]*** *Consider the solution of the Monge-Kantorovich problem in the following two cases:*

(i) μ, ν *do not give mass to small sets,* $c(x-y) = \parallel x - y \parallel^p$ *in* $\mathbb{R}^n (p > 1)$*, and the optimal transportation takes the form* $T(x) = x - \nabla\varphi(x))$*, where* φ *is given by (Theorem 8)*

(ii) μ, ν *are absolutely continuous and compactly supported in a smooth Remannian manifold M,* $c(x,y) = d(x,y)^2$*, and the optimal transportation takes the form* $T(x) = exp_x(-\nabla\psi(x))$*, where* ψ *is given by (Theorem 9) and* $\forall t \in [0,1]$ *define* T_t *by changing* ψ *for* $t\psi$ *in the expression of T. Then,* $\forall t \in [0,1], T_t$ *is also optimal in the transportation from* μ *to* $T_{\#}\mu$*.*

The proof of the theorem in [Vil03] is done in the case $c(x,y) = \| x-y \|^p$ $(p>1)$ based on a more general version of the existence theorem for optimal transport maps (8). As we did not elaborate on this it would not make much sense to give the proof here. In our case, however, there is not much to show. As $\nabla\varphi(x)$ is the gradient of a convex function, $t\nabla\varphi(x)$ is as well the gradient of a convex function. As we have seen, the transport map is uniquely determined by this gradient. Therefore the transport map between μ and $T_{t\#}\mu$ is simply T_t.
We have almost finished this chapter. I just want to review the results in words. If we view optimal transport as moving around particles, we are interested in how each of them moves from A to B. Under the aforementioned considerations the answer can be given by the solution of the time independent Monge Kantorovich problem. One just has to interpolate linearly. We have not defined the length of a curve in $\mathcal{P}(M)$ yet, but it seems reasonable to consider the family of measures obtained by our transporting process as a geodesic. The purpose of the next chapter is to provide an idea how this can be translated into a sound statement. We will, however, due to the limits of human capability just sketch the picture, and then go on in a purely formal way. Before going into this I shall mention a last very nice property obtained by our transportation process. In his paper [McC97] McCann introduced the concept of "displacement interpolation" using the fact that a absolutely continuous measure can be pushed forward onto another a.c. measure via the gradient of a convex function. This coincides with our time-dependent solution of our problem for quadratic cost functions. The family (or curve) of measures generated by this procedure

$$\rho_t = [\mu,\nu]_t \equiv [(1-t)Id + t\nabla\varphi]_{\#}\mu$$

has a remarkable property. Admitting that

$$(1-t)Id + t\nabla\varphi = \nabla[(1-t) \| \cdot \|^2 /2 + t\varphi]$$

is always the gradient of a convex function we calculate the value of the transport of $[\mu,\nu]_t$ to be

$$W_2^2(\mu,\rho_t) = \int_{\mathbb{R}^n} \| x - [(1-t)x + t\nabla\varphi(x)] \|^2 \; d\mu(x) \tag{33}$$

$$= t^2 \int_{\mathbb{R}^n} \| x - \nabla\varphi(x) \|^2 \; d\mu(x) = t^2 W_2^2(\mu,\nu) \tag{34}$$

respectively,

$$W_2(\mu,\mu_t) = tW_2(\mu,\nu). \tag{35}$$

Considering the curve generated by our transportation process (w.r.t the quadratic cost function) yields the concept of displacement convexity. Besides being a crucial tool for the investigation of functional inequalities it is a major ingredient for the theory of gradient flows on $\mathcal{P}(M)$ as developed in [AGS08]. I shall say more about this later on. Now, in the next section, we will enhance the physicist's point of view and rephrase our theory in terms of fluid mechanics.

6 Optimal Movement via Flows

Investigating the time dependent Monge Kantorovich problem in the last section, we took the trajectories of each particle into account. In the language of fluid mechanics this is called the Lagrangian point of view. This means simply that you try to understand a flow of gas or liquid by observing the path of every particle. Another approach describing the flow is to study its velocity field - the Eulerian point of view. Let us denote the family of the trajectories by $g(t,x)$, and the velocity field by $v(t, g(t,x))$. Then the two descriptions are linked together via the following equation

$$v(t, g(t,x)) = \frac{dg}{dt}(t,x).$$

We now want to analyze the Eulerian description corresponding to the family of "optimal trajectories" (T_t). Again, we are only interested in a quadratic cost function. In this case, under suitable assumptions on our initial and final measure μ and ν, such a description does exist (for each t T_t should be a diffeomorphism). Formally we ask for the evolution equation of the family of probability measures obtained via the push forward of our initial measure μ w.r.t. (T_t)

$$\rho_t = T_{t\#}\mu.$$

The following theorem is a special case of the method of characteristics. This method allows one to solve a first order PDE via a family of ODE (whose solutions might be seen as trajectories). A standard reference on the theory of PDE is [Eva98], where chapter [3.2] is devoted to the method of characteristics. Before we discuss the theorem, we give the

Definition 6.1. (Lipschitz family of diffeomorphisms; [Vil03, Definition 5.38]) A family (T_t) of mappings is said to be a locally Lipschitz family of diffeomorphisms, if

$$T_t : X \to X, \qquad \text{is bijective for all t,}$$

and for all $T < T^*$ and all compact $K \subset X$, the maps

$$(t,x) \mapsto T_t(x), \qquad \text{and } (x,t) \mapsto T_t^{-1}$$

are Lipschitz on $[0, T] \times K$.

Theorem 12. ***(Characteristics method for linear transport equations; [Vil03, Theorem 5.34])*** *Let X be $\mathbb{R}^n$, or a smooth complete manifold. Let $(T_t)_{0 \le t < T_*}$ be a locally Lipschitz family of diffeomorphisms in X, with $T_0 = Id$, and let $v = v(t, x)$ be the velocity field associated with the trajectories (T_t). Let μ be a probability measure on X, and $\rho_t = T_{t\#}\mu$. Then, $\rho_t = \rho(t, \cdot)$ is the unique solution of the linear transport equation*

$$\begin{cases} \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = 0, \quad 0 < t < T_* \\ \rho_0 = \mu \end{cases} \tag{36}$$

in $C([0, T_*); \mathcal{P}(X))$, where $\mathcal{P}(x)$ is equipped with the weak topology.

Here $\nabla\cdot$ denotes the divergence operator in the weak sense

$$\int \varphi \, d(\nabla \cdot m) = \int \nabla \varphi \cdot dm, \qquad \forall \varphi \in C_c^\infty(X).$$

The inner product $\nabla \varphi \cdot dm$ makes sense. Recall that our measure is $(v\rho)$. Thus in the weak sense the continuity equation is meant to be

$$\int_0^{T^*} \int_{\mathbb{R}^n} (\frac{\partial \varphi}{\partial t} + \nabla \varphi \cdot v) \, d\rho_t \, dt = 0, \qquad \forall \varphi \in C_c^\infty(X).$$

Now we are going to discuss the proof.

Sketch of the proof:

First we show that $\rho_t = T_{t\#}\mu$ is a solution for all t $\in (0, T)$. Recall the identities:

$$\frac{d}{dt} \int \varphi \, d\rho_t = \int (\nabla \varphi \cdot v_t) \, d\rho_t, \qquad \text{for almost all t,}$$

$$\int \varphi \, d\rho_t = \int (\varphi \circ T_t) \, d\mu,$$

and

$$\frac{\partial}{\partial t}(\varphi \circ T_t) = (\nabla \varphi \circ T_t) \cdot \frac{\partial T_t}{\partial t} = (\nabla \varphi \circ T_t) \cdot (v_t \circ T_t).$$

We now observe the "differential quotient"

$$\frac{1}{h} \left(\int \varphi \, d\rho_{t+h} - \int \varphi \, d\rho_t \right) = \int \left(\frac{\varphi \circ T_{t+h}(x) - \varphi \circ T_t(x)}{h} \right) d\mu.$$

Now, due to our assumptions on T_t and φ the expression inside the brackets converges to $(\nabla \varphi \circ T_t) \cdot (v_t \circ T_t)$ for almost all t and x. Lebesgue's dominated convergence theorem tells us that the map $t \to \int \varphi \, d\rho_t$ is differentiable for almost all t, and

$$\frac{d}{dt}\int \varphi \, d\rho_t = \int \nabla\varphi \dot{v}_t \, d\rho_t.$$

This justifies the first statement of the theorem. It remains to show the uniqueness. Because of the linearity of the equation it suffice to show that

$$\rho_0 = 0 \Rightarrow \rho_T = 0, \qquad \forall T < T^*.$$

Assume we got a Lipschitz function such that

$$\frac{\partial\varphi}{\partial_t} = -v \cdot \nabla\varphi, \qquad \varphi \mid_{t=T} = \varphi_T,$$

where φ is an arbitrary test function. Note that

$$\begin{aligned}\frac{d}{dt}\int \varphi_t \, d\rho_t &= \int \frac{\partial\varphi}{\partial t} \, d\rho_t + \int \varphi_t \, d\left(\frac{\partial\rho_t}{\partial t}\right)\\ &= -\int v_t \cdot \nabla\varphi_t \, d\rho_t + \int \varphi_t d[\nabla \cdot (v_t\rho_t)] = 0.\end{aligned}$$

This however means,

$$\int \varphi_T \, d\rho_T = \int \varphi_0 \, d\rho_0 = 0.$$

It remains to show that such a function φ exists. Such a solution is given by

$$\varphi_t(T_t s) = \varphi_T(T_T x), \qquad \varphi_t = \varphi_T \circ T_T \circ T_t^{-1}.$$

□

Thus, the velocity field, associated to our optimal transportation problem, provides a solution to the linear transport equation (or continuity equation). The next theorem will provide more information about the shape and properties of this optimal velocity field.

Theorem 13. ***(Eulerian representation for geodesic trajectories; [Vil03, Proposition 5.38]).*** *Let $v_0 : \mathbb{R}^n \to \mathbb{R}^n$ be a continusous function on $\mathbb{R}^n$, differentiable almost everywhere, and let $T_t(x) = x - tv_0(x)$ be a field of trajectories of particles, each of them moving with constant velocity. Assume that $(T_t)_{0 \leq t < T_*}$ defines a family of diffeomorphisms. Then $0 \leq t < T_*$ the associated Eulerian velocity field $v_t = T_t^{-}1 \circ dT_t/dt$ satisfies the equation*

$$\frac{\partial v}{\partial t} + v \cdot \nabla v = 0. \tag{37}$$

Sketch of the proof: As mentioned above, the Eulerian and Lagrangian point of view are linked to together via the equation

$$\frac{d}{dt}T_t(x) = v(t, T_t(x)),$$

for any x. This means

$$0 = \frac{d^2}{dt^2}(T_t x) = \frac{\partial v}{\partial t}(T_t x) + v(t, T_t x) \cdot \nabla v(t, T_T x)$$

$\square$

This yield the system of equations for our time-dependent transportation problem

$$\begin{cases} \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = 0, \quad \rho(t=0,\cdot) = \mu, \\ \frac{\partial v}{\partial t} + v \cdot \nabla v = 0. \end{cases} \tag{38}$$

So our problem is uniquely described by the above equations. Note that the cost function c does no longer appear. It is hidden in the initial conditions of the equations. More precisely

Proposition 6.2. ***(Optimal initial velocity field, [Vil03, Proposition 5.41]).*** *Assume that we are given a smooth solution of the Eulerian system* (38)*. Then, the associated Lagrangian field of trajectories determines an optimal transportation for the cost c if and only if*

$$v(t=0,\cdot) = \nabla\varphi$$

for some convex φ*.*

The goal of this section was to give a link to fluid mechanics. The purpose will become apparent in a moment.

7 The Benamou-Brenier Formula

To furnish the space of probability measures with a differentiable structure we need a Remannian metric. The theorem discussed in this chapter will give us an idea how it should look like. It will be the last theorem rigorously treated. After it our point of view will become purely formal. Maybe I should mention here that there is a well developed theory for gradient flows on $\mathcal{P}(M)$ as well, presented in [AGS08] based on the same ideas - but without the attempt to claim that $\mathcal{P}(M)$ "is" a Riemannian manifold in a common sense. I will give another remark on it later on. We will show that the "optimal" vector field we obtained by the theorems of the last section really is optimal among all reasonable vector fields. Therefore we can drop the quote signs in the sequel. In this section let ρ_0 and ρ_1 be probability densities on $\mathbb{R}^n$, and v_t some vector field moving around our particles. Let $X(t)$ denote the position of some particle at time t, then

$$\frac{d}{dt}X(t) = v_t(X(t)).$$

If our vector field is well-behaved (such that the Cauchy-Lipschitz theory applies), then we are assured of the existence of a well-defined flow on the whole time interval [0,1]. By Theorem 12 we then know, that (ρ_t) is a weak solution of the continuity equation.

$$\frac{\partial \rho_t}{\partial t} + \nabla \cdot (\rho_t v_t) = 0.$$

Again stressing the physicist's point of view, we define a total kinetic energy by

$$E(t) = \int_{\mathbb{R}^n} \rho_t(x) \parallel v_t(x) \parallel^2 dx.$$

And as usual in Newtonian mechanics an action functional

$$A[\rho, v] = \int_0^1 \left(\int \mathbb{R}^n \rho_t(x) \parallel v_t(x) \parallel^2 dx \right) dt.$$

interested in its inf. (Recall that one of the most fundamental principles in Newtonian mechanics is that a dynamical system always "wants" to move with the least possible effort). Therefore we are interested in the minimization of $A[\rho, v]$ under suitable assumptions on the density and the vector field. This finally leads us to the seminal result of Benaumou-Brenier [BB00] establishing that the minimal action in the above sense is equal to the Wasserstein distance.

Theorem 14. ***(The Benamou-Brenier formula [Vil03, Theorem 8.1])*** *Let $\rho_0, \rho_1 \in \mathcal{P}_{ac}(\mathbb{R}^n)$ be compactly supported, and $V(\rho_0, \rho_1)$ be the set of all $(\rho, v) = (\rho_t, v_t)_{0 \leq t \leq 1}$ such that*

$$\begin{cases} \rho \in C([0,1]; w* - \mathcal{P}_{ac}(\mathbb{R}^n)); \\ \\ v \in L^2(d\rho_t(x)dt); \\ \\ \bigcup_{0 \leq t \leq 1} Supp(\rho_t) \quad \textit{is bounded}; \\ \\ \frac{\partial \rho}{\partial t} + \nabla(\rho_t v_t) = 0 \quad \textit{weakly (in the distributional sense)}; \\ \\ \rho(t=0, \cdot) = \rho_0; \quad \rho(t=1, \cdot) = \rho_1 \end{cases}$$

*Here $w * -\mathcal{P}_{ac}(\mathbb{R}^n)$ is the set of absolutely continuous probability measures endowed with the weak * topology. Then*

$$\mathcal{W}_2(\rho_0, \rho_1) = \inf\{A[\rho, v]; (\rho, v) \in V(\rho_0, \rho_1)\}.$$

Sketch of the proof: The proof is given in three steps, where the second one is the technically most involved one. In the first step we show, that

$$\inf\{A[\rho, v]; (\rho, v) \in V_{sm}(\rho_0, \rho_1)\} \geq \mathcal{W}_2(\rho_0, \rho_1)$$

Where V_{sm} means that the vector field in (ρ, v) should be smooth (or more precisely bounded and C^1). We know that the Wasserstein distance is given by

$$\mathcal{W}_2(\rho_0, \rho_1) = \inf\left\{\int \rho_0(x) \mid T(x) - x \mid^2 \, dx; \; T_{\#}\rho_0 = \rho_1\right\}.$$

According to the result in the last chapter we can define the associated trajectories $T_t(x)$, and set $\rho_t = T_{t\#}\rho_0$. This means that

$$\int \rho_t(x) \mid v_t(x) \mid^2 \, dx = \int \rho_0(x) \left| \frac{d}{dt} T_t x \right|^2 \, dx.$$

Now we recall 5.1 that in the case of the quadratic cost the optimal trajectories are straight lines, and so

$$\begin{aligned} A[\rho, v] &\geq \int \rho_0(x) \left(\int_0^1 \left| \frac{d}{dt} T_t x \right|^2 \, dt \right) \, dx \\ &\geq \int \rho_0(x) \mid T_1 x - T_0 x \mid^2 \, dx, \\ &= \int \rho_0(x) \mid T_1 x - x \mid^2 \, dx. \end{aligned}$$

In the second step it is shown that the reduction to the case of smooth vector fields is justified. This is done by a change of variables. One replaces (ρ, v) by $(\rho, m) = (\rho, \rho v)$ which makes our action functional convex. Moreover one uses a mollifier to make the measures even more regular, and then shows that an approximation by smooth vector fields is sufficient. The third step guarantees the existence of a minimizing pair $(\rho, v) \in V(\rho_0, \rho_1)$. Let T $= \nabla\varphi$ be optimal, and set

$$T_t(x) = (1 - t)x + tT(x) \equiv \nabla\varphi_t(x); \qquad \rho_t = T_{t\#}\rho_0.$$

Because $\nabla\varphi_t^*$ is the inverse of $\nabla\varphi_t$ a.e. we can define a.e. the velocity field

$$v_t = \left(\frac{d}{dt} T_t\right) \circ T_t^{-1} = (T - Id) \circ T_t^{-1}.$$

It can be shown that (ρ_t, v_t) are sufficiently regular and solve the continuity equation in the weak sense. So for all nonnegative measureable function Ψ we get

$$\int \rho_t \Psi(v_t)\, dx = \int \rho_0(x) \psi(T(x) - x)\, dx.$$

Choosing $\Psi(v) = \mid v \mid^2$, we find

$$\int \rho_t(x) \mid v_t(x) \mid^2 \ dx = \int \rho(x) \mid T(x) - x \mid^2 \ dx = \mathcal{W}_2(\rho_0, \rho_1).$$

Hence

$$A[\rho, v] = \mathcal{W}_2(\rho_0, \rho_1).$$

□

This means that the Wasserstein distance w.r.t. a quadratic cost function does describe the minimal effort not just in a abstract sense but also in a physical one.

8 Intermezzo

Before introducing the heuristics we are interested in, I want to sum up the results attained so far. If one thinks about optimal transport, already with the idea that nature behaves in an optimal way in mind, one tends automatically to imagine the movement of a set of particles (e.g. a cloud of gas). To demonstrate the abstract concept of mass transportation corresponds to this physical model was the purpose of the first part. First we showed that under pretty mild conditions a solution to the problem always exits. This statement is of purely measure theoretical nature. Again I want to point out that if one would like to solve the problem in a more concrete way you would face the problem to find the solution to a PDE depending on the determinant of the Hessian of some function $f : \mathbb{R} \to \mathbb{R}^n$ (a special case of a Monge-Ampere equation), which is highly non linear, and therefore not easy to handle, and to solve this equation is a necessary condition only. Moreover, the abstract statement holds true for quiet general spaces. The next step was to link the existence theorem of an optimal transference plan to the existence of a pair of optimal functions, the Kantorovich duality. This is a very central theorem for these notes. It helped us to prove that the Wasserstein distance (equal to the value of the MKP) metrizes the weak topology of $\mathcal{P}(M)$ and to prove the existence of an optimal transport map. So we got what we wanted without treating the tedious PDE mentioned above. Another very important ingredient for the proof was the concept of Legendre transform. After we have shown (in the quadratic case only) that the pair of optimal functions in the dual problem can be restricted to a pair of convex conjugate functions we took their special structure into account. This led us to the existence of an optimal transport map and its form (recall in the case of a quadratic cost function on $\mathbb{R}^n$ it is the gradient of a convex function). Then, already heading towards a physicist's point of view, we where interested in a time-dependent version of our problem. We derived its existence based on the time-independent case (linear interpolation). This provided us with a family of trajectories and via the method of characteristics we stated the existence of a corresponding vector field solving the continuity equation. This means that we found a flow pushing our initial measure onto our final one. A very intuitive description of the problem. Then finally, the Benamou-Brenier formula showed us that the cost of our transport w.r.t. a quadratic cost function has a physical meaning as well. This we want to use to furnish $\mathcal{P}(M)$ with a Riemannian structure (only formally as already pointed out). The idea is, that our metric tensor should be given by the Benamou-Brenier formula.

Remark 8.1. In [AGS08] the correspondence between the family of densities ρ_t and the vector field v_t via the continuity equation is used to create a sound theory of gradient flows on $\mathcal{P}(M)$. They show that every absolutely continuous curve $c : [0,1] \to \mathcal{P}(M)$ corresponds to a vector field, in a reasonable way. Then you simply define the tangent space at some μ as the set consisting of all v_t related to the curves going through μ. Then they do not elaborate on the Riemannian structure of $\mathcal{P}(M)$, but concentrate on convex functional (in the sense of displacement convexity). The class of displacement convex functionals is of great interest in physics and the strategy exhibits a variety of advantages.

9 The Geometry of $\mathcal{P}(M)$

All the upcoming material is taken from [vR]. In fact it is my version of his paper. All the results are taken from there as well as the proofs are. At some points I extended them to understand them better. In this sense it will be the opposite of the last sections. The derivation of the operators needed in the sequel can be found in Lott's paper [Lot08], I will refer to it if necessary. Remember that in Riemannian geometry the central object is the metric tensor. It allows you to furnish your manifold with all the structure necessary. If you want to refresh your knowledge on (Riemannian) geometry, a standard source is [AM78]. Other good books on the subject are [Mic08], [GHL04] and [Lee03], [Jos08]. In the literature it is usually referred to Otto's paper [Ott01] as the origin of the following ideas (i.e. to treat $\mathcal{P}(M)$ as if it would be a manifold and study the behavior of certain PDE via the corresponding flow on $\mathcal{P}(M)$). In his paper he used the formal

Riemannian structure to deduce new results on the porous medium equation. **Subsequently we will concentrate on the space of probability measures of finite second moments a.c. w.r.t. the volume measure on M and smooth density functions supported on all of M. For the subsequent calculations we therefore set**

$$\mathcal{P}(M) = \left\{\mu; \int_M d^2(x,x_0)\mu\, dx < \infty;\ \frac{d\mu}{dx} \in C^\infty(M);\ \operatorname{supp}\left(\frac{d\mu}{dx}\right) = M\right\}.$$

We will often identify the measure with its density function

$$\mu \mathrel{\widehat{=}} \frac{d\mu}{dx}.$$

The idea is, that if you think about the evolution of a density μ_t, the infinitesimal variation should be a function which adds or subtracts no mass. Hence one could consider the tangent space to be

$$\mathcal{T}_\mu\mathcal{P}(M) := \left\{\psi : M \to \mathbb{R}, \qquad \int_M \psi\, dx = 0\right\}.$$

The conservation of mass is also ensured by the continuity equation

$$\frac{\partial\mu}{\partial t} = -\operatorname{div}(\mu v).$$

In our case, the vector field v should be induced by some potential φ, i.e. $v := \nabla\varphi$. Motivated by the results of the first part we make the following identification

$$\psi = -\operatorname{div}(\mu\nabla\varphi). \tag{39}$$

Note that our vector field $\nabla\varphi$ describes a curve for every $\mu \in \mathcal{P}(M)$ via its corresponding flow Φ. More precisely this means, that it induces a flow $\mu_t = (\Phi_t)_\#\mu$ on $\mathcal{P}(M)$ where Φ is the flow map induced by $\nabla\varphi$. The vector field corresponding to this flow is given by

$$V_\varphi\mu = -\operatorname{div}(\mu\nabla\varphi).$$

This means $\psi = V_\varphi\mu$ and if we assume the existence of a Green operator G_μ for $\Delta^\mu : \varphi \to -\operatorname{div}(\mu\nabla\varphi)$

$$\varphi = -G_\mu\psi \qquad \text{and } \psi = V_\varphi\mu.$$

The purpose of this identification is that we can define a reasonable norm of our tangent vectors and thus a metric tensor on $\mathcal{P}(M)$. The norm is defined to be:

$$\| \psi \|^2_{\mathcal{T}_\mu \mathcal{P}} := \int_M \| \nabla\varphi \|^2 \; d\mu.$$

In this sense the length of an optimal path between two measures μ and ν is equal to the Wasserstein distance, or the Riemannian energy of a curve $t \to \mu_t$ in $\mathcal{P}(M)$, i.e. the minimal required kinetic energy

$$E_{0,t} = \int_0^t \| \dot{\mu}_s \|^2_{\mathcal{T}_{\mu_s}\mathcal{P}(M)} \; ds = \int_0^t \int_M \| \dot{\Phi}(x,s) \|^2 \; d\mu_s ds.$$

This in fact is the result of [BB00]. Recall the the gradient of a function w.r.t a certain inner product is defined as

$$\nabla f(a) \text{ is the unique vector s.t.} \qquad df(a)h = \langle \nabla f, h \rangle.$$

Let now F be a functional and let D denote the $L^2(M, d_{\mathrm{vol}})$ Frechet derivative of F in μ. Then the "Wasserstein gradient" is computed to be

$$\nabla^{\mathcal{W}} F_{|\mu} := -\Delta^\mu (DF_{|\mu}).$$

10 Three Examples

Before we start with the investigation of the functional related to quantum mechanics, we will compute the derivatives of three other functionals.
1. Very basic $F = \int \psi(x)\mu \, dx$.

$$\begin{aligned} \frac{\partial}{\partial\varepsilon} \int \psi(x)\,(\mu + \varepsilon\eta)\,dx &= \int \psi(x)\,\eta\,dx \\ &= \langle \psi(x), \eta \rangle_{L^2(M, d_{\mathrm{vol}})} \end{aligned}$$

Thus DF $= \psi(x)$, and by identification (39)

$$\nabla^{\mathcal{W}} F_{|\mu} = \mathrm{div}(\mu \nabla \psi).$$

2. The next example is the Boltzmann entropy F $= \int \mu \, ln\,(\mu)\,dx$. Again we calculate the variation in direction η.

$$
\begin{aligned}
\frac{\partial}{\partial\varepsilon}\int(\mu+\varepsilon\eta)\cdot ln(\mu+\varepsilon\eta)\,dx &= \int \frac{(\mu+\varepsilon\eta)-\mu}{\varepsilon} ln(\mu) \\
&+ \frac{ln(\mu+\varepsilon\eta)-ln(\mu)}{\varepsilon}(\mu+\varepsilon\eta)\,dx \\
&= \int \eta\, ln(\mu) + \frac{\eta}{\mu}\mu\, dx \\
&= \int \eta\, ln(\mu) + \eta\, dx \\
&= \int (1+\ln(\mu))\,\eta\, dx \\
&= \langle 1+ln(\mu), \eta\rangle_{L^2, d_{\mathrm{vol}}}.
\end{aligned}
$$

Thus DF $= 1 + ln(\mu)$, and by identification (39)

$$
\begin{aligned}
\operatorname{div}(\mu\nabla(1+ln(\mu))) &= \operatorname{div}(\mu(0+\frac{\nabla\mu}{\mu})) \\
&= \operatorname{div}(\nabla\mu) \\
&= \Delta\mu.
\end{aligned}
$$

3. The last example is almost the functional we will use later on. It is the Fisher information functional $F = \int \parallel \nabla ln(\mu) \parallel^2 \ \mu\, dx$. To make our calculations more convenient we will compute the derivative directly (not just in direction η).

$$
\frac{\partial}{\partial\varepsilon}\int \parallel \nabla ln(\mu) \parallel^2 \ \mu\, dx = \int \parallel \nabla ln(\mu) \parallel^2 \ dx + \underbrace{\frac{\partial}{\partial\varepsilon}\int \parallel \nabla ln(\mu) \parallel^2 \ d\mu}_{*}
$$

We calculate $*$

$$
\begin{aligned}
* &= \frac{\partial}{\partial\varepsilon}\int \nabla ln(\mu)\cdot\nabla ln(\mu)\, d\mu \\
&= \int \frac{\partial}{\partial\varepsilon}\nabla ln(\mu)\cdot\nabla ln(\mu)\, d\mu + \int \nabla ln(\mu)\cdot\frac{\partial}{\partial\varepsilon}\nabla ln(\mu)\, d\mu \\
&= 2\int \frac{\partial}{\partial\varepsilon}\nabla ln(\mu)\cdot\nabla ln(\mu)\, d\mu \ = \ 2\int \nabla\frac{\partial}{\partial\varepsilon} ln(\mu)\cdot\nabla ln(\mu)\, d\mu \\
&= 2\int \left(\nabla\frac{1}{\mu}\right)\cdot\frac{1}{\mu}\nabla\mu\, d\mu \\
&= 2\int \left(\nabla\frac{1}{\mu}\right)\cdot\nabla\mu\, dx \\
&= -2\int \frac{1}{\mu}\Delta\mu\, dx.
\end{aligned}
$$

and together with the first term

$$\mathrm{DF} = \| ln(\mu) \|^2 - \frac{2}{\mu}\Delta\mu.$$

And again we identify (39)

$$\nabla \cdot (\mu\nabla(\| ln(\mu) \|^2 - \frac{2}{\mu}\Delta\mu)).$$

In physics $\| \dot{v} \|^2$ is the kinetic energy of a particle with velocity v. In terms of the above remark the fisher information functional is the kinetic energy of a measure following the heat flow. We will ad this "kinetic" term to a classical potential. This extra term will cause the "quantum effect" in our equation. What this does mean will become apparent in the sequel.

11 Schrödinger's Equation vs Madelung Equations

Now we can start with our desired investigation. In 1926, the same year as Schrödinger's work "Quantisierung als Eigenwertproblem" [Sch26] appeared, Erwin Madelung proposed an hydrodynamic interpretation "Quantenmechanik in hydrodynamischer Form" [Mad27] of Schrödinger's equation (SEQ)

$$i\hbar\partial_t\psi = -\frac{\hbar^2}{2}\Delta\psi + V\psi \tag{40}$$

This interpretation yields a system of a Hamilton-Jacobi and a continuity equation, which we will call the Madelung equations. To see this we assume that we already have a solution of the SEQ $\sqrt{\mu}e^{\frac{i}{\hbar}S}$, i.e.

$$i\hbar\partial\sqrt{\mu}e^{\frac{i}{\hbar}S} = -\frac{\hbar^2}{2}\underbrace{\Delta\sqrt{\mu}e^{\frac{i}{\hbar}S}}_{*} + V\sqrt{\mu}e^{\frac{i}{\hbar}S}$$

$$\Longleftrightarrow i\hbar\partial_t\sqrt{\mu}e^{\frac{i}{\hbar}S} + ih\sqrt{\mu}\partial_t\frac{i}{\hbar}Se^{\frac{i}{\hbar}S} = V\sqrt{\mu}e^{\frac{i}{\hbar}S} - \frac{\hbar^2}{2} * . \tag{41}$$

We now compute *.

$$\begin{aligned} * = \mathrm{div}(\nabla\sqrt{\mu}e^{\frac{i}{\hbar}S}) &= \mathrm{div}(\nabla\sqrt{\mu}e^{\frac{i}{\hbar}S} + \sqrt{\mu}\nabla e^{\frac{i}{\hbar}S}) \\ &= \mathrm{div}(\nabla\sqrt{\mu}e^{\frac{i}{\hbar}S} + \sqrt{\mu}\frac{i}{\hbar}\nabla Se^{\frac{i}{\hbar}S}) \end{aligned}$$

$$= \langle\nabla\sqrt{\mu}, \frac{i}{\hbar}\nabla Se^{\frac{i}{\hbar}S}\rangle + \Delta\sqrt{\mu}e^{\frac{i}{\hbar}S} + \langle\nabla\sqrt{\mu}, \frac{i}{\hbar}\nabla Se^{\frac{i}{\hbar}S}\rangle + \sqrt{\mu}[\langle\frac{i}{\hbar}\nabla Se^{\frac{i}{\hbar}S}, \frac{i}{\hbar}\nabla S\rangle + \frac{i}{\hbar}\Delta Se^{\frac{i}{\hbar}S}].$$

We now multiply by $\frac{1}{e^{\frac{i}{\hbar}S}}$ and get

$$
\begin{aligned}
&= \langle \nabla\sqrt{\mu}, \frac{i}{\hbar}\nabla S\rangle + \Delta\sqrt{\mu} + \langle \nabla\sqrt{\mu}, \frac{i}{\hbar}\nabla S\rangle + \sqrt{\mu}[\langle \frac{i}{\hbar}\nabla S, \frac{i}{\hbar}\nabla S\rangle + \frac{i}{\hbar}\Delta S] \\
&= \frac{2i}{\hbar}\langle \nabla\sqrt{\mu}, \nabla S\rangle + \Delta\sqrt{\mu} + \sqrt{\mu}\frac{i^2}{\hbar^2}\langle \nabla S, \nabla S\rangle + \sqrt{\mu}\frac{i}{\hbar}\Delta S \\
&= \underbrace{\frac{2i}{\hbar}\langle \nabla\sqrt{\mu}, \nabla S\rangle + \sqrt{\mu}\frac{i}{\hbar}\Delta S}_{**} + \Delta\sqrt{\mu} + \sqrt{\mu}\frac{i^2}{\hbar^2}\langle \nabla S, \nabla S\rangle
\end{aligned}
$$

$$
\begin{aligned}
** &= \frac{2i}{\hbar 2\sqrt{\mu}}\langle \nabla\mu, \nabla S\rangle + \sqrt{\mu}\frac{i}{\hbar}\Delta S \\
&= \frac{i}{\hbar\sqrt{\mu}}\operatorname{div}(\mu\nabla S)
\end{aligned}
$$

$$
\begin{aligned}
\Longrightarrow * &= \frac{i}{\hbar\sqrt{\mu}}\operatorname{div}(\mu\nabla S) + \Delta\sqrt{\mu} + \sqrt{\mu}\frac{i^2}{\hbar^2}\langle \nabla S, \nabla S\rangle \\
&= \frac{i}{\hbar\sqrt{\mu}}\operatorname{div}(\mu\nabla S) + \Delta\sqrt{\mu} - \sqrt{\mu}\frac{1}{\hbar^2} \mid \nabla S \mid^2 .
\end{aligned}
$$

Hence, the right hand side of (41) (recall that we droped $e^{\frac{i}{\hbar}S}$) has become

$$
\begin{aligned}
&V\sqrt{\mu} - \frac{\hbar^2}{2}\left[\frac{i}{\hbar\sqrt{\mu}}\operatorname{div}(\mu\nabla S) + \Delta\sqrt{\mu} - \sqrt{\mu}\frac{1}{\hbar^2} \mid \nabla S \mid^2\right] \\
&= V\sqrt{\mu} - \frac{i\hbar}{2\sqrt{\mu}}\operatorname{div}(\mu\nabla S) - \frac{\hbar^2}{2}\Delta\sqrt{\mu} + \sqrt{\mu}\frac{1}{2} \mid \nabla S \mid^2 .
\end{aligned}
$$

We now compute the left hand side of (41)

$$
i\hbar\partial_t\sqrt{\mu}e^{\frac{i}{\hbar}S} + ih\sqrt{\mu}\partial_t\frac{i}{\hbar}Se^{\frac{i}{\hbar}S}
$$

$$
= \frac{i\hbar}{2\sqrt{\mu}}\partial_t\mu e^{\frac{i}{\hbar}S} + i^2\sqrt{\mu}\partial_t Se^{\frac{i}{\hbar}S}.
$$

Again, we drop $e^{\frac{i}{\hbar}S}$ and obtain

$$
\frac{i\hbar}{2\sqrt{\mu}}\partial_t\mu - \sqrt{\mu}\partial_t S = V\sqrt{\mu} - \frac{i\hbar}{2\sqrt{\mu}}\operatorname{div}(\mu\nabla S) - \frac{\hbar^2}{2}\Delta\sqrt{\mu} + \sqrt{\mu}\frac{1}{2} \mid \nabla S \mid^2
$$

$\Longleftrightarrow$

$$
\underbrace{\frac{i\hbar}{2\sqrt{\mu}}\partial_t\mu + \frac{i\hbar}{2\sqrt{\mu}}\operatorname{div}(\mu\nabla S)}_{A} - \underbrace{\sqrt{\mu}\partial_t S - V\sqrt{\mu} + \frac{\hbar^2}{2}\Delta\mu - \sqrt{\mu}\frac{1}{2} \mid \nabla S \mid^2}_{B} = 0.
$$

Here A is the imaginary part and B the real one. As a complex number is $= 0$ iff it's imaginary and real parts are 0, we get

$$\frac{i\hbar}{2\sqrt{\mu}}\partial_t\mu + \frac{i\hbar}{2\sqrt{\mu}}\mathrm{div}(\mu\nabla S) = 0$$

$$\sqrt{\mu}\partial_t S + V\sqrt{\mu} - \frac{\hbar^2}{2}\Delta\sqrt{\mu} + \sqrt{\mu}\frac{1}{2} \mid \nabla S \mid^2 = 0$$

$$\iff \begin{cases} \partial_t\mu + \mathrm{div}(\mu\nabla S) = 0 \\ \partial_t S + \frac{1}{2} \mid \nabla S \mid^2 + V - \frac{\hbar^2}{2}\frac{\Delta\sqrt{\mu}}{\sqrt{\mu}} = 0. \end{cases} \tag{42}$$

We have shown following theorem

Theorem 15. *If the pair* (S, μ) *solves* (42), *then*

$$\psi := \sqrt{\mu}e^{\frac{i}{\hbar}S}$$

solves Schrödinger's equation

$$i\hbar\partial_t\psi = -\frac{\hbar^2}{2}\Delta\psi + V\psi$$

For further computations it will be convenient to transform the last part:

$$\frac{\hbar^2}{2}\frac{\Delta\sqrt{\mu}}{\sqrt{\mu}}.$$

We first calculate: $\Delta\sqrt{\mu}$. Set f(x):= $\sqrt{x}$

$$\begin{aligned} \Delta\sqrt{\mu} &= \Delta f(\mu) = \mathrm{div}(\nabla f(\mu)) = \mathrm{div}(f'(\mu)\nabla\mu) \\ &= f''(\mu)\underbrace{\langle\nabla\mu, \nabla\mu\rangle}_{=|\nabla\mu|^2} + f'(\mu)\Delta\mu \end{aligned}$$

We calculate $f'(\mu)$ and $f''(\mu)$

$$f'(\mu) = \frac{1}{2\sqrt{\mu}} \qquad f''(\mu) = -\frac{1}{4}\frac{1}{\mu^{\frac{3}{2}}}$$

$\Longrightarrow$

$$\begin{aligned} \frac{\Delta\mu}{\sqrt{\mu}} &= \frac{1}{\sqrt{\mu}}\left(-\frac{\mid \nabla\mu \mid^2}{4\mu^{\frac{3}{2}}} + \frac{\Delta\mu}{2\sqrt{\mu}}\right) \\ &= -\frac{\mid \nabla\mu \mid^2}{4\mu^2} + \frac{\Delta\mu}{2\mu}. \end{aligned}$$

and

$$\frac{|\nabla\mu|^2}{4\mu^2} = \frac{1}{4}\left|\frac{1}{\mu}\nabla\mu\right|^2 = \frac{1}{4}|\nabla \ln\mu|^2 .$$

Hence

$$\begin{aligned}\frac{\hbar^2}{2}\frac{\Delta\sqrt{\mu}}{\sqrt{\mu}} &= \frac{\hbar^2}{2}\left(-\frac{1}{4}|\nabla \ln\mu|^2 + \frac{\Delta\mu}{2\sqrt{\mu}}\right)\\ &= \frac{\hbar^2}{8}\left(-|\nabla \ln\mu|^2 + \frac{2}{\mu}\Delta\mu\right)\end{aligned}$$

The above computations prove the following

Corollary 11.1. *If the pair (S, μ) solves*

$$\begin{cases}\partial_t S + \frac{1}{2}\|\nabla S\|^2 + V + \frac{\hbar}{8}(\| ln(\mu)\|^2 - \frac{2}{\mu}\Delta\mu) = 0\\ \partial_t\mu + \nabla\cdot(\mu\nabla S) = 0.\end{cases} \tag{43}$$

then

$$\psi := \sqrt{\mu}e^{\frac{i}{\hbar}S}$$

solves the Schrödinger equation

$$i\hbar\partial_t\psi = -\frac{\hbar^2}{2}\Delta\psi + V\psi.$$

Now we define

$$F(\mu) := \int V(x)\,\mu\,dx + \frac{\hbar^2}{8}\int \| ln(\mu)\|^2\ \mu\,dx, \tag{44}$$

where in the second term we discover the Fisher information functional. Our first theorem states, that a solution of

$$\nabla^{\mathcal{W}}_{\dot\mu}\dot\mu = \nabla^{\mathcal{W}}F(\mu) \tag{45}$$

yields a solution of the Madelung equations, and hence a solution of the Schrödinger equation.

Remark 11.2. Recall that equation (45) is Newton's law of motion.

Theorem 16. ***[vR, Theorem 2.1]*** *Let $V \in C^\infty(M)$, and $F : \mathcal{P}^\infty(M) \to \mathbb{R}$ defined as in* (44)*. Then any smooth local solution $t \to \mu(t) \in \mathcal{P}^\infty(M)$ of* (45) *yields a local solution $(\mu_t, \overline{S}_t)$ of* (43)*, where*

$$\overline{S}(x,t) := S(x,t) + \int_0^t L_F(S_\sigma, \mu_\sigma)\, d\sigma$$

and L_F is the Lagrangian

$$L_F(\mu) := \frac{1}{2} \parallel \psi_\mu \parallel^2_{\mathcal{T}_\mu \mathcal{P}} - F(\mu) \qquad for\, \psi \in \mathcal{T}_\mu \mathcal{P}$$

and S(x,t) is the velocity potential of the flow μ, i.e. satisfying $\int S_t\, d\mu_t = 0$ and $\dot{\mu}_t = -div(\mu_t \nabla S_t)$. Conversley, let (μ_t, S_t) be a local solution of (43) *then $t \to \mu_t \in \mathcal{P}(M)$ solves* (45)*.*

Proof. Assume μ solves (45). The Wasserstein gradient $\nabla^{\mathcal{W}}$ is defined above. The calculation of the covariant derivative $\nabla^{\mathcal{W}}_{\dot{\mu}} \dot{\mu}$ can be found in [Lot08, 4.24]. Let $(x,t) \to S(x,t)$ denote the velocity potential of $\dot{\mu}$. The left-hand side of (45) is

$$\nabla^{\mathcal{W}}_{\dot{\mu}} \dot{\mu} = -\mathrm{div}\left(\mu \nabla \left(\partial_t S + \frac{1}{2} \parallel \nabla S \parallel^2\right)\right),$$

where the right-hand side is

$$\nabla^{\mathcal{W}} F = \mathrm{div}\left(\mu \nabla \left(V + \frac{\hbar^2}{8}(\parallel ln(\mu) \parallel^2 - \frac{2}{\mu}\Delta\mu)\right)\right).$$

We now define

$$Q := \partial_t S + \frac{1}{2} \parallel \nabla S \parallel^2 + V + \frac{\hbar^2}{8}(\parallel ln(\mu) \parallel^2 - \frac{2}{\mu}\Delta\mu)$$

as the sum of the terms in the inner bracket. Now

$$\mathrm{div}(\mu \nabla Q) = 0 \quad \Rightarrow \quad \int \mathrm{div}(\mu \nabla Q) \cdot Q\, dx = 0.$$

Integration by parts yields

$$\begin{aligned} \int \mathrm{div}(\mu \nabla Q) \cdot Q\, dx &= \overbrace{\oint_{\delta M} (\mu \nabla Q) \cdot Q\, dS}^{=0} - \int \mu \langle \nabla Q \cdot \nabla Q \rangle\, dx \\ &= -\int \parallel \nabla Q \parallel^2\ d\mu = 0 \end{aligned}$$

$\Longrightarrow$ (as μ is fully supported)

$$Q = const.$$

Hence

$$\partial_t S + \frac{1}{2} \parallel \nabla S \parallel^2 + V + \frac{\hbar^2}{8}(\parallel ln(\mu) \parallel^2 - \frac{2}{\mu}\Delta\mu) = c(t). \tag{46}$$

Since $\int S_t \, d\mu_t = 0$, we get

$$0 = \partial_t \langle S_t, \mu_t \rangle = \langle \partial_t S_t, \mu \rangle + \langle S_t, \partial_t \mu_t \rangle$$

and (integration by parts again),

$$\begin{aligned}
\langle S_t, \partial_t \mu_t \rangle &= \int_M S_t \; - \operatorname{div}(\mu_t \nabla S_t) \, dx \\
&= \overbrace{- \oint_{\delta M} S_t \nabla S_t \, dS}^{=0} + \int_M \nabla S_t \cdot \nabla S_t \, d\mu_t \\
&= \langle \parallel \nabla S \parallel^2, \mu_t \rangle \\
&= \overbrace{\langle c(t), \mu_t \rangle}^{=c(t)} - \frac{1}{2} \langle \parallel \nabla S \parallel^2, \mu_t \rangle - F(\mu_t) + \langle \parallel \nabla S \parallel^2, \mu_t \rangle - \overbrace{\frac{\hbar^2}{8} \langle \frac{2}{\mu} \Delta\mu, \mu_t \rangle}^{*=0} = c(t) + L_F(S_t, \mu_t).
\end{aligned}$$

To calculate * we use the divergence theorem:

$$* = \frac{2\hbar^2}{8} \int_M \frac{1}{\mu} \Delta\mu \, d\mu = \frac{\hbar^2}{4} \int_M \operatorname{div}(\nabla\mu) \, dx = \frac{\hbar^2}{4} \oint_{\delta M} \nabla\mu \, dS = 0.$$

Therefore $(\overline{S}_t, \mu_t)$ with $\overline{S}(x,t) = S(x,t) + \; L_F(S_\sigma, \mu_\sigma) \, d\sigma$ solves (43).

$$\begin{aligned}
&\partial_t \overline{S} + \frac{1}{2} \parallel \nabla \overline{S} \parallel^2 + V + \frac{\hbar^2}{8}(\parallel ln(\mu) \parallel^2 - \frac{2}{\mu}\Delta\mu) \\
&= \partial_t S + L_F + \frac{1}{2} \parallel \nabla S \parallel^2 + V + \frac{\hbar^2}{8}(\parallel ln(\mu) \parallel^2 - \frac{2}{\mu}\Delta\mu) = c(t) + L_F
\end{aligned}$$

The converse statement is now also obvious as this means that (46) is already 0.

□

We have shown that "Newton's" equation

$$\nabla^{\mathcal{W}}_{\dot{\mu}} \dot{\mu} = \nabla^{\mathcal{W}} F$$

on the Wasserstein space of probability measures yields a solution of the Schrödinger equation

$$i\hbar\partial_t\psi = -\frac{\hbar^2}{2}\Delta\psi + V\psi$$

and vice versa. This emphasizes the point of view that quantum mechanical equations are still "mechanical" equations. A fact that is less obvious in the classical formulation via the Schrödinger equation.

Remark 11.3. Note, that we should make the regularity assumptions on μ and S more precise to take advantage of the divergence theorem and integration by parts.

Remark 11.4. Note that the addition of L_F only causes an angular phase shift, thus S and $\overline{S}$ describe the same object.

In the next section we will investigate the symplectic structure of $\mathcal{TP}(M)$ induced by the Wasserstein metric tensor to affirm the equality of the standard formulation of quantum mechanics and the alternative one presented in these notes.

12 The Hamiltonian Structure

In the representation of the tangent space we will drop the notation of the foot point μ as it is given by the formula for the tangent vector anyways

$$\mathcal{TP}(M) = \{-\mathrm{div}(\mu\nabla f) \mid f \in C^\infty(M), \mu \in \mathcal{P}(M)\}.$$

Next we will investigate how our vector fields on $\mathcal{TP}(M)$ do look like. They are defined by the next

Definition 12.1. [vR, Definition 3.1] (Standard vector fields on $\mathcal{TP}(M)$). Each pair $(\psi, \varphi) \in C^\infty(M) \times C^\infty(M)$ induces a vector field $V_{\psi,\varphi}$ on $\mathcal{TP}(M)$ via

$$V_{\psi,\varphi}(-\mathrm{div}(\mu\nabla f)) = \dot{\gamma}$$

where $\gamma^{\psi,\varphi} = \gamma(t) \in \mathcal{TP}(M)$ is the curve satisfying the following properties:

$$\gamma(t) = -\mathrm{div}(\mu(t)\nabla(f + t\varphi))$$
$$\mu_t = \exp(t\nabla\psi)_\#\mu$$

Remember that the standard symplectic form on the tangent bundle of a Riemannian manifold is the exterior derivative of the canonical one form. $\omega = d\theta$. The canonical one form is defined as

Definition 12.2. (The canonical one form)

$$\theta(X) = \langle \xi, \pi_*(X)\rangle_{T_{\pi\xi}}, \quad X \in \mathcal{T}_\xi(\mathcal{T}(M)),$$

where π is the projection map $\pi : \mathcal{T}M \to M$.

We will now give a formula for the symplectic form.

Lemma 12.3. ***[vR, Proposition3.2]***
(The symplectic form on $\mathcal{TP}(M)$) *Let $\omega_{\mathcal{W}} \in \Lambda^2(\mathcal{TP}(M))$ be the standard symplectic form associated to the Wasserstein Riemannian structure on $\mathcal{P}(M)$, then*

$$\omega_{\mathcal{W}}(V_{\psi,\varphi}, V_{\tilde{\psi},\tilde{\varphi}})(-div(\mu\nabla f)) = \langle\nabla\varphi, \nabla\tilde{\psi}\rangle_\mu - \langle\nabla\tilde{\varphi}, \nabla\psi\rangle_\mu$$

Proof. Recall the identity (it can be derived via the Lie derivative)

$$\omega_{\mathcal{W}}(V_{\psi,\varphi}, V_{\tilde{\psi},\tilde{\varphi}}) = V_{\psi,\varphi}(\theta V_{\tilde{\psi},\tilde{\varphi}}) - V_{\tilde{\psi},\tilde{\varphi}}(\theta V_{\psi,\varphi}) - \theta([V_{\psi,\varphi}, V_{\tilde{\psi},\tilde{\varphi}}]),$$

where $[V_{\psi,\varphi}, V_{\tilde{\psi},\tilde{\varphi}}]$ denotes the Lie-bracket, and by the aforementioned definition

$$\theta(V_{\tilde{\psi},\tilde{\varphi}})(\text{-div}(\mu\nabla f)) = \langle\nabla f, \nabla\tilde{\psi}\rangle_\mu.$$

Now we calculate the action of the vector field $V_{\psi,\varphi}$ on the scalar valued function $\theta(V_{\tilde{\psi},\tilde{\varphi}})$, i.e.

$$\begin{aligned}
V_{\psi,\varphi}\theta(V_{\tilde{\psi},\tilde{\varphi}}) &= \frac{d}{dt}_{|t=0}\theta(V_{\tilde{\psi},\tilde{\varphi}})(\gamma^{\psi,\varphi}(t)) \\
&= \frac{d}{dt}_{|t=0}\langle\nabla(f + t\varphi), \nabla\tilde{\psi}\rangle_{\mu(t)} \\
&= \langle\nabla\varphi, \nabla\tilde{\psi}\rangle_\mu - \int_M \nabla f \cdot \nabla\tilde{\psi}(\text{-div}(\mu\nabla\psi))\, dx \\
&= \langle\nabla\varphi, \nabla\tilde{\psi}\rangle_\mu + \int_M \nabla(\nabla f \cdot \nabla\tilde{\psi})\nabla\psi\, d\mu,
\end{aligned}$$

where here in the last step we used the divergence theorem and in the line above integration by parts. As our canonical one form θ only measures the projections, we get

$$\theta([V_{\psi,\varphi}, V_{\tilde{\psi},\tilde{\varphi}}])(\text{-div}(\mu\nabla f)) = \langle\nabla f, [\nabla\psi, \nabla\tilde{\psi}]\rangle_\mu.$$

Collecting the terms we get:

$$\langle\nabla\varphi, \nabla\tilde{\psi}\rangle_\mu - \langle\nabla\tilde{\varphi}, \nabla\psi\rangle_\mu + \int_M \nabla(\nabla f\cdot\nabla\tilde{\psi})\nabla\psi\, d\mu - \int_M \nabla(\nabla f\cdot\nabla\psi)\nabla\tilde{\psi}\, d\mu - \langle\nabla f, [\nabla\psi, \nabla\tilde{\psi}]\rangle_\mu.$$

We show that

$$\int_M \nabla(\nabla f \cdot \nabla\tilde{\psi})\nabla\psi\, d\mu - \int_M \nabla(\nabla f \cdot \nabla\psi)\nabla\tilde{\psi}\, d\mu - \langle\nabla f, [\nabla\psi, \nabla\tilde{\psi}]\rangle_\mu = 0$$

which concludes the proof.

$$\int_M \nabla(\nabla f \cdot \nabla\tilde{\psi})\nabla\psi \, d\mu - \int_M \nabla(\nabla f \cdot \nabla\psi)\nabla\tilde{\psi} \, d\mu - \langle \nabla f, \nabla_{\nabla\psi}\tilde{\psi}\rangle_\mu + \langle \nabla f, \nabla_{\nabla\tilde{\psi}}\psi\rangle$$

$$\underbrace{\int_M \nabla(\nabla f \cdot \nabla\tilde{\psi})\nabla\psi \, d\mu - \langle \nabla f, \nabla_{\nabla\psi}\tilde{\psi}\rangle}_{*} - \int_M [\nabla(\nabla f \cdot \nabla\psi)\nabla\tilde{\psi} \, d\mu - \langle \nabla f, \nabla_{\nabla\tilde{\psi}}\psi\rangle_\mu]$$

Here * is just the formula for the Hessian $Hess(\tilde{\psi}, \psi)$. As it is symmetric this means that the second term $Hess(\psi, \tilde{\psi})$ is equal to the first, hence the difference is 0. To see that this term is the Hessian we recall that it is given by

$$\nabla df(X, Y) = \langle \nabla_X \nabla f, Y\rangle.$$

Now we calculate

$$\begin{aligned}
\nabla(\nabla f \cdot \nabla\tilde{\psi})\nabla\psi &= \langle \nabla\langle \nabla f, \nabla\tilde{\psi}\rangle, \nabla\psi\rangle \\
&= \nabla_{\nabla\psi}\langle \nabla f, \nabla\tilde{\psi}\rangle \\
&= \langle \nabla_{\nabla\psi}\nabla f, \nabla\tilde{\psi}\ \rangle + \langle \nabla f, \nabla_{\nabla\psi}\tilde{\psi}\rangle
\end{aligned}$$

Here we got in the left hand side the definition of the Hessian plus a term. The same term is subtracted in *, hence * indeed is $Hess(\tilde{\psi}, \psi)$.

$\square$

Using the Riemannian inner product in each fiber of $\mathcal{TP}(M)$ the Hamiltonian associated to L_F is

$$H_F : \mathcal{TP}(M) \to \mathbb{R}; \quad H_F(\text{-div}(\mu\nabla f)) = \frac{1}{2}\int_M \| \nabla f \|^2 \ d\mu + F(\mu)$$

Now we are prepared to calculate the Hamiltonian vector field.

Lemma 12.4. ***[vR, Proposition 3.4]*** *Let X_F denote the Hamiltonian vector field induced on $\mathcal{TP}(M)$ from H_F and $\omega_{\mathcal{W}}$, then*

$$X_F(\text{-}div(\mu\nabla f)) = V_{f, -(\frac{1}{2}\|\nabla f\|^2 + V + \frac{\hbar^2}{8}(\|ln\mu\|^2 - 2\frac{\Delta\mu}{\mu}))}(\text{-}div(\mu\nabla f)).$$

Proof. Fix $\psi, \varphi \in C^\infty(M)$ and let $V_{\psi,\varphi(.)}$ denote the corresponding standard vector field, and $\gamma(t)$ the associated curve as defined above, then

$$V_{\psi,\varphi}(H_F)(\text{-div}(\mu\nabla f)) = \frac{d}{dt}_{|t=0}(H_F\gamma(t))$$

$$= \frac{d}{dt}_{|t=0}\left(\overbrace{\frac{1}{2}\int \| \nabla(f+t\varphi) \|^2 \; d\mu_t}^{I} \;\; + \;\; \overbrace{\langle V, \mu_t\rangle}^{II} \;\; + \;\; \overbrace{\frac{\hbar^2}{8}I(\mu_t)}^{III}\right)$$

We calculate I, II and III seperatly

$$\begin{aligned} I &= \frac{d}{dt}_{|t=0}\frac{1}{2}\int \| \nabla(f+t\varphi) \|^2 \; d\mu_t \\ &= \frac{1}{2}\int \frac{d}{dt}_{|t=0}[\| \nabla(f+t\varphi) \|^2]\, d\mu + \frac{1}{2}\int \| \nabla f \|^2 \, \dot{\mu}\, dx \\ &= \langle \nabla f, \nabla\varphi\rangle_\mu + \frac{1}{2}\int \| \nabla f \|^2 \, (\text{-div}(\mu\nabla f))\, dx \\ &= \langle \nabla f, \nabla\varphi\rangle_\mu + \langle \nabla\psi, \nabla(\frac{1}{2} \| \nabla f \|^2)\rangle_\mu . \end{aligned}$$

where we have used integration by parts to get equality between the last two lines.

$$II = \frac{d}{dt}_{|t=0}\int V\, d\mu_t$$

$$= \int V\dot{\mu}\, dx$$

$$\begin{aligned} &= \int V(\text{-div}(\mu\nabla f)) \\ &= \int \nabla V \cdot \nabla\psi\, d\mu \\ &= \langle \nabla V, \nabla\psi\rangle_\mu . \end{aligned}$$

And finally the third term,

$$\begin{aligned} III &= \frac{d}{dt}_{|t=0}\frac{\hbar^2}{8}\int \| \nabla ln\, \mu \|^2 \; d\mu \\ &= \frac{\hbar^2}{8}\int 2\nabla ln\, \mu \nabla\left(\frac{\text{-div}(\mu\nabla\psi)}{\mu}\right) d\mu + \frac{\hbar^2}{8}\int \| \nabla ln\, \mu \|^2 \, (\text{-div}(\mu\nabla psi))\, d\mu \\ &= \frac{\hbar^2}{8}\left(\langle \nabla\psi, \nabla\left(-\frac{2}{\mu}\Delta\mu\right)\rangle + \langle \nabla\psi, \nabla \| \nabla ln\, \mu \|^2\rangle\right). \end{aligned}$$

The last term is just the derivative of the Fisher information functional in direction $\nabla\psi$ at μ. Recall, that our μ_t is the geodesic in direction $\nabla\psi$ on $\mathcal{P}(M)$. Summing up I, II and III this yields

$$V_{\psi,\varphi}(H_F)(\text{-div}(\mu\nabla f)) =$$

$$= \langle\nabla f, \nabla\varphi\rangle_\mu - \langle\nabla(-(\frac{1}{2}\parallel\nabla f\parallel^2 + V + \frac{\hbar^2}{8}(\parallel\nabla ln\,\mu\parallel^2 - \frac{2}{\mu}\Delta\mu))), \nabla\psi\rangle_\mu.$$

□

Remark 12.5. As expected, the last theorem proves that the integral curves of the Hamiltonian vector field w.r.t our Riemannian metric and modified potential correspond to the solutions of the Madelung flow.

13 Equivalence via a Symplectic Submersion

This is the final section on the alternative representation of the Schrödinger equation. It relates the symplectic structure presented in the previous section to the "standard" symplectic structure on $\mathcal{C} = C^\infty(M;\mathbb{C})$ the space of smooth complex valued functions on M.

Definition 13.1. [vR, Definition 4.1] (Symplectic subersion). A smooth map $s : (M,\omega) \to (N,\eta)$ between two symplectic manifolds is called a symplectic submersion if its differential $s_* : \mathcal{T}M \to \mathcal{T}N$ is surjective and satisfies $\eta(s_*X, s_*Y) = \omega(X,Y)$ for all $X, Y \in \mathcal{T}M$.

The next proposition tells us how Hamiltonian flows are transformed under symplectic submersions.

Proposition 13.2. ***[vR, Proposition 4.2]*** *Let $s : (M,\omega) \to (N,\eta)$ be a symplectic submersion and let $f \in C^\infty(M)$ and $g \in C^\infty(N)$ with $g \circ s = f$, then s maps Hamiltonian flows associated to f on (M,ω) to Hamiltonian flows associated to g on (N,η).*

The solutions of the Schrödinger equation belong to the set $\mathcal{C} = C^\infty(M;\mathbb{C})$, the space of complex valued smooth functions on M. We identify the tangent space $\mathcal{TC}$ over an element $\psi \in \mathcal{C}$ with $\mathcal{C}$, where $\mathcal{TC}$ is equipped with the symplectic form

$$\omega_{\mathcal{C}}(F,G) = -2\int Im(F\cdot\overline{G})(x)\,dx.$$

The Schrödinger equation as a Hamiltonian flow on $\mathcal{C}$ is induced from the symplectic form $\hbar\cdot\omega_C$ and the Hamiltonian function

$$H_S(\psi) = \frac{\hbar^2}{2} \int \parallel \nabla\psi \parallel^2 \; dx + \int \parallel \psi(x) \parallel^2 V(x)\, dx.$$

We now shrink the set $\mathcal{C}$ to the subset of note $\mathcal{C}_*$, the set of nowhere vanishing functions on $\mathcal{C}$ such that $\int \parallel \psi(x) \parallel^2 \; dx = 1$. This set is invariant under the Schrödinger flow. Under the assumption that M is simply connected (and a theorem of algebraic topology) there exists a "polar-like" decomposition of each $\psi \in \mathcal{C}_*$, i.e. $\psi = \parallel \psi \parallel e^{\frac{i}{\hbar}S}$ where $S : M \to \mathbb{R}$ smooth is uniquely defined up to an additive constant $\hbar 2\pi k$, $k \in \mathbb{N}$. Motivated by our above computations we define the Madelung transform to be

$$\sigma : \mathcal{C}_*(M) \to \mathcal{TP}(M), \quad \sigma(\psi) = \text{-div}(\parallel \psi \parallel^2 \nabla S).$$

Now the final statement establishes that the Madelung transform is a symplectic submersion between the two structures of interest.

Theorem 17. ***[vR, Theorem 4.3]*** *Let M be simply connected. Then the Madelung transform*

$$\sigma : \mathcal{C}_*(M) \to \mathcal{TP}(M), \quad \sigma(\parallel \psi \parallel e^{\frac{i}{\hbar}S}) = \text{-}div(\parallel \psi \parallel^2 \nabla S)$$

defines a symplectic submersion from $(\mathcal{C}_*(M), \hbar \cdot \omega_{\mathcal{C}})$ to $(\mathcal{TP}(M), \omega_{\mathcal{W}})$ which preserves the Hamiltonian, i.e.

$$H_S = H_F \circ \sigma.$$

Proof. First we note that $\sigma(\mathcal{C}_*(M)) = \mathcal{TP}(M)$ (clearly $\parallel \psi \parallel^2$ is a measure and ∇S belongs via our identification to the tangent space at $\parallel \psi \parallel^2$). We now show that $\sigma(\mathcal{C}_*(M)) = \mathcal{TP}(M)$ is a submersion. Therefore we fix a point $x_0 \in M$, then for each $r \in [0, 2\pi\hbar)$, let $\tau = \tau^{(r)}$

$$\tau : \mathcal{TP}(M) \to \mathcal{C}_*, \quad \text{-div}(\mu\nabla S) \to \sqrt{\mu} e^{\frac{i}{\hbar}(S-(S(x_0)-r))}.$$

τ is a bijection from $\mathcal{TP}(M)$ to $\{\psi \in \mathcal{C}_*, \frac{\psi}{\|\psi\|}(x_0) = e^{\frac{i}{\hbar}r}\}$, and satisfies $\sigma \circ \tau = Id_{\mathcal{TP}(M)}$. This verifies the surjectivity of the differential σ_* of σ. Now we are going to show that σ preserves the symplectic structure. Therefore we set $\psi = e^{\frac{i}{\hbar}f} \in \mathcal{C}_*$ with $f(x_0) = r \in [0, 2\pi\hbar)$ and let $\eta = \text{-div}(\mu\nabla f) = \sigma(\psi) \in \mathcal{TP}(M)$. As we already have shown that $\sigma \circ \tau = Id_{\mathcal{TP}(M)}$ it remains to proof that $\tau^*\omega_C = \frac{1}{\hbar} \cdot \omega_{\mathcal{W}}$ on $\mathcal{T}_\eta(\mathcal{TP}(M))$. Moreover, as we know that the set $\{V_{\psi,\varphi}(\text{-div}(\mu\nabla f) \mid \psi, \varphi \in C^\infty(M)\}$ spans the full tangent space $\mathcal{T}_\eta(\mathcal{TP}(M))$, we can restrict ourself to establish the identity

$$\omega_C(\tau_* V_{\psi,\varphi}, \tau_* V_{\tilde{\psi},\tilde{\varphi}}) = \frac{1}{\hbar}\omega_{\mathcal{W}}(V_{\psi,\varphi}, V_{\tilde{\psi},\tilde{\varphi}})$$

$\forall \psi, \varphi, \tilde{\psi}, \tilde{\varphi} \in C^\infty(M)$. By definition of $V_{\psi,\varphi}$ and $\tau = \tau^{(r)}$ for $\mu_t := \exp(t\nabla\psi)_*(\mu)$ and $c(t) := f(x_0) + t\varphi(x_0) - r$, it follows that

$$\tau_* V_{\psi,\varphi} = \frac{d}{dt}\Big|_{t=0} \sqrt{\mu_t} e^{\frac{i}{\hbar}(f+t\varphi - c(t))} = e^{\frac{i}{\hbar}f}\left(\frac{1}{2\sqrt{\mu}}(\text{-div}(\mu\nabla\psi)) + \sqrt{\mu}\frac{i}{\hbar}(\varphi - \dot{c})\right).$$

Hence

$$\begin{aligned}\omega_C(\tau_* V_{\psi,\varphi}, \tau_* V_{\tilde{\psi},\tilde{\varphi}}) &= -2\int \frac{1}{2\sqrt{\mu}}(\text{-div}(\mu\nabla\psi)) \cdot -\sqrt{\mu}\frac{1}{\hbar}(\tilde{\varphi} - \dot{c}) \\ &\quad + \int \sqrt{\mu}\frac{1}{\hbar}(\varphi - \dot{c}) \cdot (\text{-div}(\mu\nabla\tilde{\psi})) \\ &= \frac{1}{\hbar}(\langle\nabla\psi, \nabla\tilde{\varphi}\rangle_\mu - \langle\nabla\varphi, \nabla\tilde{\psi}\rangle_\mu) = \frac{1}{\hbar}\omega_{\mathcal{W}}(V_{\psi,\varphi}, V_{\tilde{\psi},\tilde{\varphi}}).\end{aligned}$$

It remains to show that $H_S = H_F \circ \sigma$. Let $\psi = \tau(\text{-div}(\mu\nabla f))$, then $\nabla\psi = \sqrt{\mu} e^{\frac{i}{\hbar}f(\frac{1}{2}\nabla \ln \mu + \frac{i}{\hbar}\nabla f)}$, and remember that $H_F(\text{-div}(\mu\nabla f)) = \frac{1}{2}\int \parallel f \parallel^2 d\mu + F(\mu)$. Then we see that

$$\frac{\hbar^2}{2}\int \parallel \nabla\psi \parallel^2 = \frac{1}{2}\int \parallel \nabla f \parallel^2 \; d\mu + \frac{\hbar^2}{8} I(\mu).$$

And $\int \parallel \psi \parallel^2 \; V(x)\, dx = \langle V, \mu\rangle$. Adding this term at each side of the equation shows that $H_S = H_F \circ \sigma$, and thus concludes the proof.

□

As we already pointed out, our description of the motion of a quantum particle corresponds to Newton's equation on the Wasserstein space. This section showed that there is a symplectic lifting to the higher dimensional space $\mathcal{C}_*(M)$. The lifted equation is Schrödinger's equation which is linear, and therefore much easier to handle. Mapping the solution back to the Wasserstein space via σ requires a correction in the phase field. We already mentioned above that this does not affect the object but just its representation.

14 Final Remarks

All the calculations can easily be performed for the non-linear Schrödinger equation

$$i\partial_t\psi = -\frac{1}{2}\Delta\psi + \kappa \mid \psi \mid^2 \psi$$

as well, replacing the potential V $= \int V(x)\, d\mu$ by $\frac{1}{2}\int \kappa\mu\, d\mu$. Here $\hbar$ is set to be 1 as the nonlinear Schrödinger equation describes classical phenomena in optics and the theory of water waves. Another interesting case is the Schrödinger equation in a magnetic field

$$i\hbar\partial_t\psi = -\frac{\hbar^2}{2}\Delta\psi + \frac{i\hbar e}{c}A\cdot\nabla\psi + \frac{e^2}{c^2}A^2\psi + e\varphi\psi.$$

This should lead to a modification of the metric on the Wasserstein space. Till now, however, it is unknown how the equation can be rephrased in a Newtonian form on the space $\mathcal{P}(M)$. The idea to use the fisher information functional can be traced back to the paper [HR02] of Hall and Reginatto. Their interpretation is a statistical one, though. In statistics the fisher information is a tool for optimal parameter estimation. However, if one already does know some facts about optimal transport the interpretation of the fisher information functional as a kind of energy is not made up out of thin air. An other paper on Hamilton-Jacobi equations and optimal transport is [GNT08].

References

[AGS08] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[AM78] Ralph Abraham and Jerrold E. Marsden. *Foundations of mechanics*. Benjamin/Cummings Publishing Co. Inc. Advanced Book Program, Reading, Mass., 1978. Second edition, revised and enlarged, With the assistance of Tudor Raţiu and Richard Cushman.

[BB00] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84(3):375–393, 2000.

[Bil99] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[EG92] Lawrence C. Evans and Ronald F. Gariepy. *Measure theory and fine properties of functions*. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992.

[Els05] Jürgen Elstrodt. *Maß- und Integrationstheorie*. Springer-Lehrbuch. [Springer Textbook]. Springer-Verlag, Berlin, fourth edition, 2005. Grundwissen Mathematik. [Basic Knowledge in Mathematics].

[Eva98] Lawrence C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1998.

[GHL04] Sylvestre Gallot, Dominique Hulin, and Jacques Lafontaine. *Riemannian geometry*. Universitext. Springer-Verlag, Berlin, third edition, 2004.

[GNT08] Wilfrid Gangbo, Truyen Nguyen, and Adrian Tudorascu. Hamilton-Jacobi equations in the Wasserstein space. *Methods Appl. Anal.*, 15(2):155–183, 2008.

[HR02] Michael J. W. Hall and Marcel Reginatto. Schrödinger equation from an exact uncertainty principle. *J. Phys. A*, 35(14):3289–3303, 2002.

[Jos08] Jürgen Jost. *Riemannian geometry and geometric analysis*. Universitext. Springer-Verlag, Berlin, fifth edition, 2008.

[KN76] John L. Kelley and Isaac Namioka. *Linear topological spaces.* Springer-Verlag, New York, 1976. With the collaboration of W. F. Donoghue, Jr., Kenneth R. Lucas, B. J. Pettis, Ebbe Thue Poulsen, G. Baley Price, Wendy Robertson, W. R. Scott, and Kennan T. Smith, Second corrected printing, Graduate Texts in Mathematics, No. 36.

[Lee03] John M. Lee. *Introduction to smooth manifolds*, volume 218 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, 2003.

[Lot08] John Lott. Some geometric calculations on Wasserstein space. *Comm. Math. Phys.*, 277(2):423–437, 2008.

[Mad27] Erwin Madelung. Quantentheorie in hydrodynamischer form. *Zeitschrift fr Physik*, 40:322–326, 1927.

[McC97] Robert J. McCann. A convexity principle for interacting gases. *Adv. Math.*, 128(1):153–179, 1997.

[McC01] Robert J. McCann. Polar factorization of maps on Riemannian manifolds. *Geom. Funct. Anal.*, 11(3):589–608, 2001.

[Mic01] Radu Miculescu. Approximation of continuous functions by Lipschitz functions. *Real Anal. Exchange*, 26(1):449–452, 2000/01.

[Mic08] Peter W. Michor. *Topics in differential geometry*, volume 93 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 2008.

[Ott01] Felix Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.

[Roc97] R. Tyrrell Rockafellar. *Convex analysis.* Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1997. Reprint of the 1970 original, Princeton Paperbacks.

[Sch26] Erwin Schrdinger. Quantisierung als eigenwertproblem. *Ann. Phys*, 79:361–376, 1926.

[Vil03] Cédric Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 2003.

[Vil09] Cédric Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, 2009. Old and new.

[vR] Max von Renesse. An optimal transport view on schroedinger's equation. *to appear in Canad. Math. Bull.*

[Wer00] Dirk Werner. *Funktionalanalysis.* Springer-Verlag, Berlin, extended edition, 2000.

Abstract

Zusammenfassung

In der vorliegenden Arbeit "Schrödinger's Equation as Newtons's Law of Motion" werden neuere Ergebnisse aus dem Gebiet der Transporttheorie vorgestellt und gezeigt wie die wohl bekannteste Gleichung der Quantenmechanik, die Schrödingergleichung als Newtonsche Gleichung geschrieben werden kann. Also als eine klassische Bewegungsgleichung.
Zu Beginn werden die fr die Arbeit wichtigsten Ergebnisse der Transporttheorie behandelt. Die Existenz einer optimalen Transportstrategie im Sinne von Kantorovich steht ganz zu Beginn. Es folgen die Definition einer Metrik auf dem Raum der Wahrscheinlichkeitsmaße und wichtige Eigenschaften dieser Metrik (z.B.: Metrisierung der schwachen Topologie).
Danach wird aus der Existenz einer Lsung des Transportproblems im Sinne von Kantorovich die Existenz einer optimalen Transportabbildung hergeleitet. Eine solche Transportabbildung wird auch Lsung des Mongeproblems genannt. Die Herleitung dieser optimalen Abbildung erfolgt nur fr den speziellen Fall der fr die Arbeit wichtigen Kostenfunktion (die Distanzfunktion des Grundraums zum Quadrat).
Mit Hilfe dieser Abbildung werden dann die wichtigsten Resultate fr die angestrebte Geometrisierung des Raumes der Wahrscheinlichkeitsmae gezeigt. Zuerst wird mittels der Transportabbildung eine ganze Familie von Transportabbildungen definiert. Diese Familie von Abbildungen erlaubt es dann Geodäten auf dem Raum der Wahrscheinlichkeitsmae zu definieren.
In Abschnitt 6 werden diesen Geodten Vektorfelder zugeordnet und mittels dieser Vektorfelder auch Geschwindigkeitsvektoren.
In Abschnitt 7 wird dann mit Hilfe der Benamou-Brenier-Formel eine Norm (bzw. inneres Produkt) von Geschwindigkeitsvektoren definiert. Dieses innere Produkt ist von zentralem Interesse. Mit Hilfe dieses Produktes wird dann (formal) eine differenzierbare Struktur, im Sinne der Riemannschen Geometrie, auf dem Raum der Wahrscheinlichkeitsmae definiert.
Die Notation fr diese Geometrie wird in Abschnitt 9 festgelegt.
Danach folgen einige Beispiele. Es werden die Gradienten verschiedener Funktionale berechnet. Eines dieser Beispiele (die Fisherinformation) ist auch fr den weiteren Teil der Arbeit von hchstem Interesse.
Um eine Verbindung der Schrdinger Gleichung zu der von uns angestrebten Newtonschen Gleichung herzustellen ist ein Zwischenschritt erforderlich. In Abschnitt 11 wird ein auf Erwin Madelung zurckgehendes Resultat vorgestellt. Es zeigt, dass die Schrdinger Gleichung in ein System von Gleichungen (eine Hamiltion -Jacobi und eine Kontinuittsgleichung) umgeschrieben werden

kann.
Die zentrale Aussage erfolgt dann in Theorem 16. Es wird gezeigt, dass eine Newtonsche Gleichung auf deren rechten Seite ein spezielles Potential steht zu den Madelunggleichungen quivalent ist. Das angesprochene Potential besteht aus einem klassischen Potential und der sogenannten Fisherinformation.
Im letzten Teil der Arbeit wird die quivalenz der Gleichungen mittels der Methoden der symplektischen Geometrie gezeigt, um die Aussage in aller Allgemeinheit darzustellen.

Lebenslauf

Persönliche Daten

Name:	Fuchs
Vorname:	Philipp
Geburtsdatum:	24.08.1982
Adresse:	Gentzgasse 115/1/8; 1180 Wien
Telefonnummer:	0699 81 15 09 76

Schulbildung

September 1996 bis Juni 2001:	Handelsakademie

Zivildienst

Oktober 2001 bis Oktober 2002	Krankenhaus St. Josef in Braunau

Studienverlauf

Oktober 2002 bis April 2003:	Studium Physik und technische Mathematik an der Universität Innbruck
Oktober 2003 bis Mai 2010:	Studium der Mathematik an der Universitt Wien

Auslandsaufenthalte

1 Semester ERASMUS in Paris (Sommersemester 2007)
3 Wochen Summer School zu Optimal Transport in Grenoble (Juni 2009)
2 Monate in Berlin im Rahmen der Diplomarbeit (Januar, Februar 2010)